



Faculteit Psychologie en  
Pedagogische Wetenschappen

**Reading strategies instruction and peer tutoring in primary schools.  
A quasi-experimental study of the effects on students'  
reading comprehension and metacognition**

Hilde Van Keer

Promotor: Prof. Dr. J. P. Verhaeghe

Proefschrift ingediend tot het behalen van de academische graad  
van Doctor in de Pedagogische Wetenschappen

2002

# Contents

## Voorwoord

## Chapter 1

<b>General introduction</b>	1
Outline of Theoretical and Empirical Background	1
Research Questions	5
Operationalization and Measurement Instruments	5
Design of the Studies	7
Structure of the Dissertation	8
References	11

## Chapter 2

<b>Effects on second and fifth graders' reading comprehension and self-efficacy related thoughts: Results of the first study</b>	17
Abstract	17
Introduction	17
Method	21
Results	31
Discussion	39
References	44

## Chapter 3

<b>Reanalyzing the first study's effects on fifth graders' reading comprehension: A three-level repeated measures approach</b>	51
Abstract	51
Introduction	51
Method	57
Hypotheses	66
Results	67
Discussion	77
References	81

## Chapter 4

<b>Effects of peer tutoring on second graders' reading fluency</b>	89
Abstract	89
Introduction	89
Method	91
Results	96
Discussion	100
References	102

<b>Chapter 5</b>	
<b>Comparing two teacher training programs for innovating reading comprehension instruction</b>	<b>107</b>
Abstract	107
Introduction	107
Description of the Teacher Training Strategies	108
Study's Purpose and Hypothesis	111
Method	112
Results	117
Discussion	122
References	126
 <b>Chapter 6</b>	
<b>Effects on second and fifth graders' reading comprehension, strategy use and self-efficacy related thoughts: Results of the second study</b>	<b>131</b>
Abstract	131
Introduction	131
Study's Purpose and Hypotheses	134
Description of the Innovative Approach	135
Method	138
Results	144
Discussion	151
References	157
 <b>Chapter 7</b>	
<b>General discussion</b>	<b>163</b>
Overview of the Results with regard to Student Outcomes	164
General Discussion	176
Limitations of the Present Studies and Suggestions for Subsequent Research	185
Conclusion	190
References	192

## Voorwoord

Aan het tot stand komen van dit proefschrift hebben heel wat mensen bijgedragen. Iedereen die me bij het opzetten en uitvoeren van het onderzoek of bij de rapportage ervan met raad en daad heeft bijgestaan, verdient hier dan ook een woordje van dank. Allereerst dank ik graag mijn promotor Jean Pierre Verhaeghe, alsook de overige leden van de doctoraatsbegeleidingscommissie Wilfried De Corte, Lieven Verschaffel en Lut Van Damme voor hun ondersteuning, kritische blik en nuttige commentaren. Daarnaast denk ik aan de mensen in het veld die actief betrokken waren bij het onderzoek: de honderden "leesmakkers" van de tweede en vijfde klas die met niet aflatend enthousiasme participeerden en alle schoolteams en leraren waarmee ik een aantal schooljaren lang nauw samenwerkte. Hun inzet, reflecties, vragen en boeiende bedenkingen getuigden van een sterke betrokkenheid en waren een permanente aansporing om door te gaan. Verder kon ik ook steeds terugvallen op mijn collega's van de Vakgroep Onderwijskunde. Met hen van gedachten wisselen was steeds inspirerend en verhelderend. Tot slot richt ik mij nog tot alle anderen die elk op hun eigen manier hun steentje hebben bijgedragen aan het tot stand komen van dit proefschrift. Dank jullie wel, voor de onmisbare steun en stimulans, voor de hulp in moeilijke momenten, de nodige duwtjes in de rug en voor nog zo veel meer.

Gent, 9 september 2002.

# Chapter 1

## General introduction

### Outline of Theoretical and Empirical Background

It is widely recognized that learning to read is one of the most crucial learning processes children are involved in at primary school. Becoming a proficient reader is, however, not that easy for everyone. Especially with regard to the ultimate goal of reading comprehension many children appear to have huge and persistent problems. Moreover, prevailing reading instruction seems not able to handle students' comprehension difficulties adequately.

Reading comprehension can be defined as constructing a mental representation of textual information and its interpretation (Van Den Broek & Kremer, 2000) and refers to understanding and putting a meaning on written words, sentences, and texts (Aarnoutse & Van Leeuwe, 1998, 2000). Notwithstanding the apparent simplicity of the definition, there are complex processes behind it.

Previously, reading comprehension was assumed to occur automatically once students could decode correctly (Dole, 2000). The conceptual shift to a cognitive model of learning in the seventies, however, forced researchers to rethink the comprehension process and to study the underlying cognitive processes involved in reading comprehension. Cognitively based views of reading comprehension emphasize that expert readers distinguish themselves from novice or poor readers by the flexible use of a repertoire of comprehension-monitoring and regulating activities (Baker & Brown, 1984; Dole, Duffy, Roehler, & Pearson, 1991). This implies metacognition, which refers to the knowledge or awareness of someone's cognitive possibilities and activities and to the ability or skills to manage and control these cognitive processes and evaluate whether or not they are performing successfully (Baker & Brown, 1984; Gersten, Fuchs, Williams, & Baker, 2001; Gourgey, 2001). In other words, proficient readers have a split focus: they are able to concentrate simultaneously on the material and on themselves, reflecting on what they are doing while they read. They are aware of whether they understand or do not understand, and this awareness usually leads to regulation and repair when they encounter difficulties (Dole et al., 1991). To that end, they have a variety of appropriate strategies at their disposal, which are thought of as conscious, instantiated, and flexible plans selected deliberately, deployed, and adapted to a variety of texts, purposes, and occasions (Baker & Brown, 1984; Paris, Wasik, & Turner, 1991; Pressley & Allington, 1999; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989).

The mastery of cognitive and metacognitive strategies appears to be critical in becoming a skilled reader. However, not all primary students spontaneously discover and develop both the metacognitive skills to monitor and regulate comprehension and the "fix up" cognitive strategies to repair understanding when it breaks down (Hartman, 2001; Pressley & Allington, 1999). Despite the convincing research results that explicit teacher-mediated strategies instruction can effectively promote students'

strategic reading and reading comprehension (e.g., Brand-Gruwel, Aarnoutse, & Van Den Bos, 1998; De Corte, Verschaffel, & Van De Ven, 2001; Dole et al., 1991; Duffy et al., 1987; Haller, Child, & Walberg, 1988; Klingner & Vaughn, 1996; Klingner, Vaughn, & Schumm, 1998; Palincsar & Brown, 1984; Pressley, 2000; Pressley et al., 1989; Walraven, 1995), there is evidence that little has changed since Durkin's (1978-1979) now classic observation research into reading comprehension instruction. The dominant instructional practice of teaching reading is hardly characterized by explicit and continuous comprehension instruction aimed at making students metacognitively astute in strategy selection, use, and evaluation, but focuses mainly on testing and evaluating students' understanding by questioning them about the content of a text after reading it (Aarnoutse, 1995; Aarnoutse & Weterings, 1995; Dole, 2000; Paris & Oka, 1986; Pressley, Wharton-McDonald, Hampston, & Echevarria, 1998; Weterings & Aarnoutse, 1986). Moreover, Aarnoutse (1995) has even argued that, notwithstanding the fact that contemporary Dutch reading methods take into account the importance of strategies, the practice of reading instruction still remains very traditional.

Theories about metacognitive reading strategies should be embedded within a more general theoretical framework about metacognition. Metacognition can be considered as a multidimensional construct. As to its components, a distinction can be made between metacognitive knowledge or knowledge of cognition and metacognitive skills or regulation of cognition (Gourgey, 2001; Schraw, 2001; Veenman & van Hout-Wolters, 2002). The former refers to what individuals know about their own cognition or cognition in general. More particularly, it includes three kinds of metacognitive awareness: declarative, procedural, and conditional knowledge. Declarative knowledge includes knowledge about oneself as a learner and about the factors influencing one's performance. Procedural knowledge refers to knowledge about how to do things. Conditional knowledge refers to knowing when and why to use declarative and procedural knowledge (Brown, 1987; Cross & Paris, 1988; Garner, 1990). The second component, regulation of cognition, refers to a set of activities that help students to manage and control their cognitive processes. Generally, three metacognitive control processes are distinguished, namely planning, monitoring, and evaluation (Schraw, 2001). Finally, some researchers (Flavell, 1979; Simons, 1996) also distinguish metacognitive beliefs as a third separate component of metacognition that refers to an individual's opinions with regard to someone's cognitive abilities, processes, and factors that can influence these, such as attributions of success and failure and self-efficacy perceptions. Others, however, argue that beliefs about the self are better seen as motivational constructs (Pintrich, Wolters, & Baxter, 2000). Whatever conceptualization is used, it is generally agreed that students' beliefs about their abilities are important for learning.

In addition to the mastery of cognitive and metacognitive strategies, learning to read appears to be a social process, in which the interaction between the reader and the writer, as well as the interaction among readers occupy an important position.

Although there is general consensus on the importance of communication between learners for the development of knowledge and understanding and on the effectiveness of various forms of peer learning, the underlying mechanisms responsible for the

benefits are often approached and elucidated from different theoretical perspectives. Piaget stressed the significance of interaction between peers because it helps children to "decenter" and to become sensitive to other perspectives on the world than their own. Recent researchers in the Piagetian tradition use the concept of socio-cognitive conflict to take account of how a child's understanding may be shifted by interacting with another child who has a rather different understanding of events. The basic idea is that when two contrasting world views are brought into contact, this is likely to stimulate some cognitive restructuring, learning, and improved understanding (Mercer, 1996). From this perspective it can be argued that peers facilitate meaningful interpretation of texts for each other because in peer-led interaction and discussions cognitive conflicts arise, inadequate reasoning is exposed, and cognitive restructuring and higher-quality understanding emerges (Alvermann, 2000; Slavin, 1995, in Gambrell, Mazzoni, & Almasi, 2000). On the other hand, Vygotsky's sociocultural theory and some neo-Vygotskians differ from the Piagetians by emphasizing cooperation rather than conflict, suggesting that learners having to explain ideas to each other is useful because it encourages the development of a more explicit, organized, distanced kind of understanding (Mercer, 1996). More specifically, Vygotsky (1978) emphasized the transition from interpersonal to intrapersonal functioning and argued in this respect that social interaction may act as a steppingstone and even be a prerequisite for independent developmental achievement. What learners learn to do today in interaction, they will internalize and do independently tomorrow. With regard to reading comprehension instruction this implies that peer-led interaction and collaborative literacy experiences provide opportunities for students to actively and substantively engage in the co-construction of meaning as they share and exchange ideas about texts (Gambrell et al., 2000). The result of participating in such a social situation involving reading and thinking about texts is that individual students can draw upon other students to help them construct not only an understanding of text ideas and content but also an understanding of the process of constructing meaning from text, in other words of what it means to read and think about text (Kucan & Beck, 1997).

Notwithstanding the different theoretical approaches, these theories are not incompatible with one another. Each focuses on a different aspect of the interactive and collaborative process. Sociocultural theory, for example, may offer insight into the co-construction of meaning, but the concept of cognitive conflict must be borrowed from Piaget as a mechanism to explain individual cognitive change (Van Meter & Stevens, 2000).

Research evidence convincingly establishes that opportunities to participate in social interaction among peers and collaborative settings make up an important part of reading instruction that aims at an actual increase in comprehension, higher level-cognition and the application of self-regulation strategies (e.g., Almasi, 1996; Almasi, McKeown, & Beck, 1996; Brown, Pressley, Van Meter, & Schuder, 1996; Dole et al., 1991; Fuchs & Fuchs, 2000; Fuchs, Fuchs, Mathes, & Simmons, 1997; Greenwood, Delquadri, & Hall, 1989; Guthrie, Schafer, Wang, & Afflerbach, 1993; Klingner & Vaughn, 1996; Klingner et al., 1998; Mathes & Fuchs, 1994; Palincsar & Brown, 1984; Pressley et al., 1992; Rosenshine & Meister, 1994; Simmons, Fuchs, Fuchs,

Mathes, & Hodge, 1995). Given these persuasive findings, one might expect that teachers allot a fair amount of the reading comprehension instruction periods to peer interaction and discussion. Nevertheless, student-centered discussion is anything but common practice in most classrooms (Alvermann, 2000). Researchers have, however, explored a number of procedures to encourage social and task-oriented interaction among peers in the classroom. In this respect "peer tutoring" appears to be a promising instructional technique.

Peers helping one another in situations of learning is not a new experience. As Topping (1996) describes, peer tutoring is an old instructional practice, traceable back as far as Antiquity. In those days, the peer tutor was more specifically perceived as a surrogate teacher in a linear model of the transmission of knowledge from teacher to tutor to tutee. Since that time peer tutoring was repeatedly rediscovered as an effective instructional technique. In the late eighteenth and early nineteenth century, for example, the pioneers Andrew Bell and Joseph Lancaster developed a tutoring or monitoring system in India and England to face the problem of the large number of students and the shortage of schooled teachers (Fuchs et al., 1997). In the late 1960s again a rejuvenation of interest occurred. This was more specifically based on the opportunities peer tutoring creates for offering individualized and intensive instruction to academically needy students (Gerber & Kauffman, 1981). As development and research into peer tutoring proceeded, it became clear that peer tutoring is not necessarily only about transmission from the more able and experienced to the less able (Topping, 1996). The tutor as well seems to benefit from the peer tutoring experience. Therefore, peer tutoring at present can be defined as "people from similar social groupings who are not professional teachers helping each other to learn, and learning themselves by teaching" (Topping, 1996). The term peer tutoring more specifically covers a whole series of practices that employ peers as one-to-one teachers to provide individualized instruction, practice, repetition and clarification of concepts (Topping, 1988; Utley & Mortweet, 1997) and is characterized by specific role taking: at any point someone has the job of tutor, while the other is in a role as tutee (Topping, 1996).

Besides typologies of peer tutoring in terms of offered structure, intensity, curriculum content, role continuity, or objectives, two large categories of peer tutoring can be distinguished on the basis of the composition of the dyads. Children can be paired with other children from within their own classroom. This variant is called same-age tutoring. The specific form of same-age tutoring in which the students alternate on a regular basis between the tutor and tutee role is called reciprocal same-age tutoring (Fantuzzo, King, & Heller, 1992). The second variant is called cross-age tutoring and refers to older students tutoring younger students. In this case role reciprocity is basically not applicable.

Within the framework of the striking contrast between instructional practice and research evidence establishing the importance of explicit reading strategies instruction and peer-led interaction about texts, the main objective of this dissertation is to design, implement, and evaluate an innovative instructional approach, aimed at making reading comprehension instruction in primary schools more effective by systematically



blending research-based practices from both research fields. More specifically, the experimental treatment that was designed comprises two innovative cornerstones, namely explicit instruction in relevant reading strategies and opportunities to practice strategic reading by stimulating interaction between peers about texts in either cross-age or reciprocal same-age peer tutoring dyads. In addition, two distinct teacher training programs will be compared on their efficacy & strategies to translate the innovations into classroom practice.

Before proceeding with reporting the studies and research results, the main research questions, the operationalization and applied measurement instruments, the design of the studies, and the structure of the dissertation will be presented.

## **Research Questions**

The dissertation contains the reportage of the results from two successive studies, in which the effectiveness of the innovative approach to reading comprehension instruction was examined through the implementation in complete regular second- and fifth-grade classes during an entire school year. More specifically, both studies jointly focus on the following research questions:

1. What influence does explicit reading strategies instruction, followed by practice in (a) teacher-led whole-class activities, (b) reciprocal same-age, or (c) cross-age peer tutoring activities, have on primary school students' reading skills and social and emotional development?
2. Are the effects of explicit reading strategies instruction followed by practice in (a) reciprocal same-age or (b) cross-age peer tutoring dyads different in comparison with identical instruction followed by practice in teacher-led whole-class activities? More specifically, do the peer tutoring sessions in which the functional application of reading strategies is practiced have a surplus value with respect to the teacher-led practice of these strategies?
3. Are the effects of explicit strategies instruction followed by practice in cross-age peer tutoring dyads different in comparison with identical instruction followed by practice in reciprocal same-age dyads?
4. What are the effects of a restricted in-service teacher training compared to a year-round elaborated coaching of teachers with a view to implement the innovative approach to reading comprehension instruction on teachers' experiences with the attended training course and on students' progress in reading skills and social and emotional development?

## **Operationalization and Measurement Instruments**

To assess students' reading skills, data were collected with regard to second and fifth graders' strategic reading behavior and reading comprehension achievement. In addition, second-grade data were supplemented with reading fluency scores. As

concerns students' reading behavior, we confined ourselves to students' self-reports of the application of both cognitive and metacognitive reading strategies. To that end, a questionnaire was developed. To estimate students' reading comprehension and fluency achievement standardized tests were used (Brus, 1969; Staphorsius & Krom, 1996; Verhoeven, 1993). The choice of the comprehension tests was based on their well-established psychometric characteristics, the built-in adaptation to different student abilities and the fact that the tests address aspects of reading comprehension that are covered by the reading strategies, which are part of the experimental program. In this respect the reading comprehension tests used can be considered as "sensitive" to the kind of learning that was aimed at in the experimental treatments.

As concerns the emotional development, we focused on assessing students' metacognitive beliefs with regard to their reading proficiency. In this respect, a self-report questionnaire was designed to measure the incidence of students' thoughts related to self-efficacy perceptions with regard to reading and accompanying success and failure attributions. It should be noted that capturing the incidence of self-efficacy related thoughts does not give a direct measure of students' self-efficacy perception, but rather an indication of the degree to which a student is preoccupied with such kind of thoughts. In that sense the data are more directly related to (meta)cognitive activity than data collected by means of more traditional self-concept questionnaires. However, a high incidence of negative self-efficacy related thoughts can surely be considered as an indication of a low self-efficacy perception, but such a conclusion can not be drawn from a low incidence of positive self-efficacy related thoughts.

Although we mainly focused on students' self-efficacy perceptions that are directly related to reading activities, an existing self-concept questionnaire (Veerman, Straathof, Treffers, Van den Bergh, & ten Brink, 1997), which is a Dutch version of the Self-Perception Profile for Children (Harter, 1985), was administered as well. Because this questionnaire was not suitable for second graders, the instrument was solely used with the fifth-grade age group.

As regards social development, we focused on estimating students' sociometric rating score (Singleton & Asher, 1977). More specifically, the sociometric position of students can be considered as a reliable and valid measure of the quality of students' social relations and acceptance within the class group (Van Oost, Braem, De Ruyck, & Mommerency, 1989; Whitley, Schofield, & Snyder, 1984) on the one hand and as an indicator of students' social competence and interaction skills (Asher & Hymel, 1981; Asher & Renshaw, 1981; Coie & Dodge, 1988; Denham, McKinley, Couchoud, & Holt, 1990; Erwin, 1993; Putallaz & Gottman, 1981; Rubin & Daniels-Beirness, 1983) on the other hand.

Finally, students' attitude towards reading was included as an additional student outcome variable, which can be considered as a mediating process factor with a possible impact on students' reading skills and thoughts relating to self-efficacy perceptions. In this respect a Dutch questionnaire was used to collect the data (Aarnoutse, 1996).

With regard to teachers' experiences with the attended training course, a questionnaire was developed to assess teachers' satisfaction with the training, perception of workload with regard to implementing the innovative instructional approach, and estimated student progress as a result of the implementation of the innovations.

## **Design of the Studies**

As mentioned above, the dissertation contains the reportage of the results from two successive studies. The first study is characterized by a quasi-experimental research with a pretest posttest retention test control group design with in total 444 second-grade and 454 fifth-grade students from 44 classes in 25 different schools throughout Flanders.

In order to examine the research questions, participating teachers and their classes were assigned to four different research conditions. In the strategies-only condition the experimental treatments included explicit instruction in reading strategies, followed by practicing the strategies in teacher-led whole-class settings. The experimental same-age and cross-age peer tutoring condition included identical explicit instruction in the same reading strategies, but strategic reading was respectively practiced in reciprocal same-age and cross-age dyads. Finally there was a control group, characterized by traditional reading comprehension instruction without explicit instruction or peer tutoring. It should be mentioned that all activities that were part of the treatment in the experimental conditions were conducted by the regular classroom teachers and embedded in regular school time during periods normally allocated for reading instruction. The interventions were not meant as an additional program on top of the teachers' traditional reading comprehension classes. Experimental teachers were coached to substitute their traditional way of teaching comprehension with one of the experimental programs.

The second study comprises two parts. On the one hand, this study is a partial replication of the first study. Again, a quasi-experimental pretest posttest retention test control group design was used. Participants included 396 second graders and 449 fifth graders from 42 different classrooms. More particularly, two experimental peer tutoring and one control condition was distinguished. The experimental same-age and cross-age peer tutoring condition included explicit reading strategies instruction, followed by practicing the application of the strategies in respectively reciprocal same-age and cross-age dyads. Based on the experiences from the first study the program and some of the materials were slightly modified.

On the other hand, the second study compares the efficacy of two different teacher training courses as strategies to support teachers in developing competence for implementing the innovative approach. The 14 second- and 16 fifth-grade teachers who were involved in one of the experimental conditions received either an intensive year-round coaching or a more restricted 13-hours in-service teacher training. Most of the teachers involved in the second study had also participated in the first study. Teachers from the first study's cross-age or same-age peer tutoring condition continued with the same experimental treatment and got further intensive coaching. Teachers

who had previously been involved in the strategies-only condition were assigned to the restricted in-service training program, in which they were introduced and supported to implement either the cross-age peer tutoring treatment or the same-age peer tutoring treatment. Since in some schools parallel classes were created as a result of an expanding number of second- or fifth-grade students in comparison with the first study, a few new teachers joined the study. They were assigned to the same experimental treatments as their colleagues from the same school. Taking into account that almost all teachers participated for the second school year, in comparing both teacher training courses one has to bear in mind that these teachers were already experienced in implementing at least part of the experimental treatments. Nevertheless, teachers in the extensive coaching condition began training with a full year of experience in implementing peer tutoring, while this was completely new for teachers in the restricted in-service condition. Moreover, the former teachers already received a full school year of intensive coaching during the first study, whereas teachers in the restricted training only obtained more limited information and support about instructing reading strategies. In this respect, we have to call attention to the fact that potential differences in the effects of the training conditions are probably not only attributable to differences in the experimental treatment, but also to the different starting point of the teachers.

## Structure of the Dissertation

The following chapters dilate upon the theoretical rationale of the research, go into a full consideration of the innovative instructional approach designed within the framework of this dissertation, and chronologically discuss the results of the successive analyses of the data collected in each study. More specifically, they deal with the research results revealing significant effects of the experimental treatments. Chapters 2 to 6 have been submitted for publication in international journals. Therefore, they can be read separately.

Before briefly presenting an overview of the chapters, Table 1.1 gives an overall picture of both studies' research questions that are discussed in the different chapters.

Table 1.1  
*Overview of the Studies' Research Questions Addressed in the Different Chapters*

	Study 1	Study 2	Integration
Chapter 1			
Chapter 2 <sup>a</sup>	Research question 1, 2, 3		
Chapter 3 <sup>b</sup>	Research question 1, 2, 3		
Chapter 4 <sup>b</sup>	Research question 1, 2, 3		
Chapter 5 <sup>b</sup>		Research question 4	
Chapter 6 <sup>b</sup>		Research question 1 (b, c), 3	
Chapter 7			Research question 1, 2, 3, 4

<sup>a</sup> Currently under revision after review by an international journal.

<sup>b</sup> Submitted for publication in an international journal.

The first study addresses the first three research questions. Chapter 2 reports the research results with regard to the effects of the experimental interventions on both second and fifth graders' reading comprehension achievement and self-efficacy related thoughts. To take the hierarchical nesting of students into a smaller number of classes into account, the effectiveness of the interventions is analyzed by applying multilevel modeling. More specifically, two levels are distinguished, namely students (level 1) clustered within classes (level 2), and separate hierarchical regression analyses are executed on students' post- and retention test data, adjusted for their pretest scores and other possible explanatory factors before experimental condition was entered in the equation.

Bearing in mind that multilevel models are considered very useful for analyzing longitudinal data that are collected at fixed or varying occasions (Snijders & Bosker, 1999), Chapter 3 explores the feasibility of an alternative analysis of the data, better adjusted to the repeated measures design of the study. More specifically, a special kind of hierarchical nesting can be defined with regard to the three waves of student data (Goldstein, 1995; Jones & Duncan, 1998; Paterson & Goldstein, 1991; Rasbash et al., 1999; Snijders & Bosker, 1999), leading to a three-level structure with the level-one measurement occasions clustered within the level-two students, who are in their turn nested within level-three classes. In Chapter 3 the fitness of the three-level model was solely tested with regard to fifth graders' reading comprehension achievement. Just as in the two-level analyses possible explanatory factors were entered in the equations before experimental condition was included. It was found that the three-level structure is an appropriate and more specialized technique to analyze the longitudinal data collected in the study. Therefore, this hierarchical three-level repeated measures model was applied further in subsequent analyses, reported from Chapter 4 onwards. The application and construction of appropriate multilevel models was disclosed step by step in both Chapter 2 and 3. The results sections in the subsequent chapters are somewhat less elaborated, but can be understood according to similar lines of reasoning.

In Chapter 4, the focus is on a side effect of the innovative instructional approach on students' reading fluency achievement. Notwithstanding the fact that the core of the experimental interventions centers on reading comprehension, based on the available literature it was hypothesized that the considerable amounts of active and individualized reading practice in regularly organized peer tutoring reading activities would promote fluent reading as well (Delquadri, Greenwood, Whorton, Carta, & Hall, 1986; Fuchs et al., 1997; Greenwood et al., 1989; Simmons, Fuchs, Fuchs, Hodge, & Mathes, 1994; Simmons, Fuchs, Fuchs, Mathes, & Pate, 1995). Chapter 4 more specifically discusses second graders' progress in reading fluency. No reading fluency data were collected in the fifth-grade classes, since it is recognized that reading fluency is generally well developed at the end of grade 3 (Bast & Reitsma, 1998; Sticht & James, 1984).

As mentioned above, the second study comprises two parts. One part addresses the fourth general research question. The other part of the second study is a partial replication of the first study and deals with research question 1 (subdivision b and c) and 3.

Chapter 5 presents the results with regard to the differences between the intensive year-round coaching condition and the restricted in-service training condition. The differential effects of both teacher training strategies on teachers' experiences with the attended training course and perceptions of treatment effects on the students were investigated by means of multivariate and univariate analyses of variance. The differential effects on students' progress in reading skills and social and emotional development were investigated using hierarchical three-level repeated measures analyses. Both conditions were founded on research-based effective teacher training components (Djalil & Anderson, 1989; Harchik, Sherman, Sheldon, & Strouse, 1992; Showers, 1990; Veenman, Van Tulder, & Voeten, 1994; Wheldall, Merrett, & Borg, 1985), but distinguished themselves by the intensity of support lent to the teachers. The results of the first study confirmed the positive impact of a year-round support. However, the extensive assistance provided to the teachers was considered as a restriction of the study, for it is unlikely that regular in-service training would be able to offer commensurable levels of support. Therefore, this part of the second study was aimed at developing an equally effective, but less intensive teacher training, geared to the levels of support that are achievable for training authorities. Taking into account the built-in common research-based training features, it was hypothesized that teachers' experiences as well as student outcomes would be comparable in both training courses. In the case of differential effects, however, the training program teachers' attended should be included as an explanatory variable in the subsequent analyses that investigate the effectiveness of the implementation of the innovative approach in the classrooms on student outcomes in the replication study. However, the results reported in Chapter 5 revealed no significant differential effects of the teacher training programs on student outcomes, as a result of which the analyses with regard to the replication study could be executed on the combined data of both teacher training conditions.

The results of the latter analyses are presented in Chapter 6. More specifically, the chapter discusses the research results with regard to the effects of the experimental interventions on both second and fifth graders' reading comprehension achievement, reports of reading strategy use, and self-efficacy related thoughts in the second study. The analyses were performed applying hierarchical three-level repeated measures models.

Before passing to a general discussion of the research findings, the results of both studies with regard to student outcomes are assembled. In this respect, one has to bear in mind that two different analysis techniques were applied to the data. Initially, in Chapter 2, two-level hierarchical regression analyses were applied separately on students' post- and retention test data. Subsequently, the analysis technique switched to the use of three-level repeated measures models from Chapter 3 onwards. Since for the first study three-level analyses were solely executed with regard to fifth graders' reading comprehension achievement (Chapter 3) and second graders' reading fluency (Chapter 4), the comparison of the research results from the first and second study can not completely be drawn correctly. To obviate this restriction, the first study's data that were solely analyzed by means of the two-level models were reanalyzed by means of hierarchical repeated measures models as applied to the data of the second study. The

review of the results is reported in Chapter 7. Following this review, a general discussion of the results of both studies is presented. Finally, Chapter 7 concludes with an outline of the limitations of the studies, and suggestions for future research.

## References

- Aarnoutse, C. A. J. (1995). *Probleemidentificatie en aanzet voor een actieplan taal* [Problem identification and initiative towards an action plan for language]. Nijmegen, The Netherlands: Vakgroep Onderwijskunde.
- Aarnoutse, C. A. J. (1996). *Leesattitudeschaal* [Reading attitude scale]. Nijmegen, The Netherlands: Berkhout.
- Aarnoutse, C. A. J., & Van Leeuwe, J. F. J. (1998). Relation between reading comprehension, vocabulary, reading pleasure, and reading frequency. *Educational Research and Evaluation*, 4, 143-166.
- Aarnoutse, C. A. J., & Van Leeuwe, J. F. J. (2000). Development of poor and better readers during the primary school. *Educational Research and Evaluation*, 6, 251-278.
- Aarnoutse, C. A. J., & Weterings, A. C. E. M. (1995). Onderwijs in begrijpend lezen [Education in reading comprehension]. *Pedagogische Studiën*, 72, 82-101.
- Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.
- Almasi, J. F., McKeown, M. G., & Beck, I. L. (1996). The nature of engaged reading in classroom discussions of literature. *Journal of Literacy Research*, 28, 107-146.
- Alvermann, D. E. (2000). Classroom talk about texts: Is it dear, cheap or a bargain at any price? In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 170-192). New York: Teachers College Press.
- Asher, S. R., & Hymel, S. (1981). Children's social competence in peer relations: Sociometric and behavioral assessment. In J. D. Wine, & M. D. Smye (Eds.), *Social competence* (pp. 125-157). New York: The Guilford Press.
- Asher, S. R., & Renshaw, P. D. (1981). Children without friends: Social knowledge and social-skill training. In S. R. Asher, & J. M. Gottman (Eds.), *The development of children's friendships* (pp. 273-296). Cambridge: Cambridge University Press.
- Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353-394). New York: Longman.
- Bast, J. W., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, 34, 1373-1399.
- Brand-Gruwel, S., Aarnoutse, C. A. J., & Van Den Bos, K. P. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and instruction*, 8, 63-81.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other mysterious mechanisms. In F. Weinert, & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Erlbaum.

- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.
- Brus, B. Th. (1969). *Eén-minuut-test* [One minute test]. Nijmegen, The Netherlands: Berkhout.
- Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development*, 59, 815-829.
- Cross, D. R., & Paris, S. G. (1988). Development and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology*, 80, 131-142.
- De Corte, E., Verschaffel, L., & Van De Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Educational Psychology*, 71, 531-559.
- Delquadri, J. C., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional Children*, 52, 535-542.
- Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of preschool peer ratings. *Child Development*, 61, 1145-1152.
- Djalil, A., & Anderson, L. W. (1989). The impact of a research-based teacher training program on Indonesian teachers, classrooms and students. *Teaching and Teacher Education*, 5, 165-178.
- Dole, J. A. (2000). Explicit and implicit instruction in comprehension. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 52-69). New York: Teachers College Press.
- Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61, 239-264.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliff, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putnam, J., & Bassari, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22, 347-368.
- Durkin, D. (1978-1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14, 481-533.
- Erwin, Ph. (1993). *Friendships and peer relations in children*. Chichester: John Wiley & Sons.
- Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology*, 84, 331-339.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.
- Fuchs, L. S., & Fuchs, D. (2000). Building student capacity to work productively during peer-assisted reading activities. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 95-115). New York: Teachers College Press.



- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Gambrell, L. B., Mazzoni, S. A., & Almasi, J. F. (2000). Promoting collaboration, social interaction and engagement with text. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging young readers. Promoting achievement and motivation* (pp. 119-140). New York: The Guilford Press.
- Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research*, 60, 517-529.
- Gerber, M. M., & Kauffman, J. M. (1981). Peer tutoring in academic settings. In P. Strain (Ed.), *The utilization of classroom peers as behavior change agents* (pp. 155-188). New York: Plenum.
- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, 71, 279-320.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Gourgey, A. F. (2001). Metacognition in basic skills instruction. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 33-68). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.
- Guthrie, J. T., Schafer, W., Wang, Y. Y., Afflerbach, P. (1993). Relationships of instruction to amount of reading: An exploration of social, cognitive, and instructional connections. *Reading Research Quarterly*, 30, 8-25.
- Haller, E. P., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher*, 17, 5-8.
- Harchik, A. E., Sherman, J. A., Sheldon, J.B., & Strouse, M.C. (1992). Ongoing consultation as a method of improving performance of staff members in a group home. *Journal of Applied Behavior Analysis*, 25, 599-610.
- Harter, S. (1985). *Manual for the self-perception profile for children*. Denver: University of Denver.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 33-68). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Jones, K., & Duncan, C. (1998). Modeling context and heterogeneity: Applying multilevel models. In E. Scarbrough, & E. Tanenbaum (Eds.), *Research strategies in the social sciences. A guide to new approaches* (pp.95-123). Oxford, United Kingdom: Oxford University Press.
- Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Primary School Journal*, 96, 275-293.
- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Primary School Journal*, 99, 3-21.

- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271-299.
- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.
- Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction*, 6, 359-377.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition and motivation. *Developmental Review*, 6, 25-56.
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Vol. 2* (p. 609-640). New York: Longman.
- Paterson, L., & Goldstein, H. (1991). New statistical methods for analyzing social structures: An introduction to multilevel models. *British Educational Research Journal*, 17, 387-393.
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw, & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43-97). Lincoln, NE: Buros Institute of Mental Measurements.
- Pressley, M. (2000). Comprehension instruction in primary school: A quarter-century of research progress. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 32-51). New York: Teachers College Press.
- Pressley, M., & Allington, R. (1999). What should reading instructional research be the research of? *Issues in Education*, 5, 1-35.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Primary School Journal*, 92, 513-555.
- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Primary School Journal*, 90, 3-32.
- Pressley, M., Wharton-McDonald, R., Hampston, J. M., & Echevarria, M. (1998). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies of Reading*, 2, 159-191.
- Putallaz, M., & Gottman, J. M. (1981). Social skills and group acceptance. In S. R. Asher, & J. M. Gottman (Eds.), *The development of children's friendships* (pp. 116-149). Cambridge: Cambridge University Press.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, 64, 479-530.

- Rubin, K. H., Daniels-Beirness, T. (1983). Concurrent and predictive correlates of sociometric status in kindergarten and grade 1 children. *Merrill-Palmer Quarterly*, 29, 337-351.
- Schraw, G. (2001). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 3-16). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Showers, B. (1990). Aiming for superior classroom instruction for all children: A comprehensive staff development model. *Remedial and Special Education*, 11, 35-39.
- Simmons, D. C., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice*, 9, 203-212.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Primary School Journal*, 95, 387-408.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., & Pate, H. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Simons, P. R. (1996). Metacognition. In E. De Corte & F. E. Weinert (Eds.), *International encyclopedia of developmental and instructional psychology* (pp.441-445). New York: Pergamon.
- Singleton, L. C., & Asher, S. R. (1977). Peer preferences and social interaction among third-grade children in an integrated school district. *Journal of Educational Psychology*, 69, 330-336.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced modeling*. London: Sage.
- Staphorsius, G., & Krom, R. (1996). *Toetsen Begrijpend Lezen* [Reading Comprehension Tests]. Arnhem, The Netherlands: Cito.
- Sticht, T. G., & James, J. H. (1984). Listening and reading. In P. D. Pearson (Ed.), *Handbook of reading research* (Vol. 1, pp. 293-318). New York: Longman.
- Topping, K. J. (1988). *The peer tutoring handbook: Promoting cooperative learning*. London: Croom Helm.
- Topping, K. J. (1996) The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.
- Utley, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children*, 29, 69-92.
- Van Den Broek, P., & Kremer, K. E. (2000). The mind in action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 1-31). New York: Teachers College Press.
- Van Meter, P., & Stevens, R. J. (2000). The role of theory in the study of peer collaboration. *Journal of Experimental Education*, 69, 113-127.

- Van Oost, P., Braem, M., De Ruyck, H., & Mommerency, G. (1989). *Bevorderen van sociale competentie bij kinderen. Van onderzoek naar praktijk in de klas*. [Promoting social competence for children. From research to educational practice]. Leuven, Belgium: Acco.
- Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2002). *Het meten van metacognitive vaardigheden* [Measuring metacognitive skills]. Paper presented at the ORD-conference, May 30<sup>th</sup>, 2002, Antwerp.
- Veenman, S., Van Tulder, M., & Voeten, M. (1994). The impact of in-service training on teacher behavior. *Teaching and Teacher Education*, 10, 303-317.
- Veerman, J. W., Straathof, M. A. E., Treffers, Ph. D. A., Van den Bergh, B. R. H., & ten Brink, L. T. (1997). *Competentiebelevingsschaal voor Kinderen (CBSK)* [Self-Perception Profile for Children]. Lisse, The Netherlands: Swets & Zeitlinger.
- Verhoeven, L. (1993). *Lezen met begrip 1* [Reading with comprehension 1]. Arnhem, The Netherlands: Cito.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walraven, M. (1995). *Instructie in leesstrategieën. Problemen met begrijpend lezen en het effect van instructie aan zwakke lezers* [Instruction in reading strategies. Difficulties with reading comprehension and the effect of instruction for poor readers]. Amsterdam: Paedologisch Instituut.
- Weterings, A. C. E. M., & Aarnoutse, C. A. J. (1986). De praktijk van het onderwijs in begrijpend lezen [Instruction in reading comprehension in practice]. *Pedagogische Studiën*, 63, 387-400.
- Wheldall, K., Merrett, F., & Borg, M. (1985). The behavioural approach to teaching package (BATPACK): An experimental evaluation. *British Journal of Educational Psychology*, 55, 65-75.
- Whitley, B. E. Jr., Schofield, J. W., & Snyder, H. N. (1984). Peer preferences in a desegregated school: A round robin analysis. *Journal of Personality and Social Psychology*, 46, 799-810.

## Chapter 2

### Effects on second and fifth graders' reading comprehension and self-efficacy related thoughts: Results of the first study\*

#### Abstract

The present study evaluated the effectiveness of explicit reading comprehension strategies instruction, followed by practice in teacher-led whole-class activities (STRAT), reciprocal same-age (STRAT+SA), or cross-age peer tutoring activities (STRAT+CA) on second and fifth graders' reading comprehension and self-efficacy perceptions. For second grade multilevel analyses revealed significant STRAT and STRAT+CA effects that however did not last after finishing the program. Fifth-grade students in all three experimental conditions performed significantly better on the posttest than their control group peers. A continued growth till at least six months after finishing the program was shown for the STRAT and STRAT+CA conditions. Moreover, on both post- and retention test fifth graders in the STRAT+CA condition reported significantly less negative thoughts related to their reading proficiency.

#### Introduction

A frequent misconception about students' reading skills is that if children are able to decode the words, they understand the meaning of the written information. Such a misconception confuses the skill of decoding with comprehension (Ezell & Kohler, 1992). While decoding instruction has had a long history of attention and debate, comprehension instruction has only recently received more attention.

Previously, reading comprehension was considered to be a process of mastering: once students could decode, comprehension was assumed to occur automatically (Dole, 2000). Research however shows that good readers are characterized by more than just decoding. Cognitively based views of reading comprehension emphasize that proficient readers use a flexible repertoire of comprehension-monitoring and regulating activities (Dole, Duffy, Roehler, & Pearson, 1991). They have a purpose for reading, they get acquainted with the text, monitor comprehension as they read, reflect on the reading process, and evaluate their understanding. Moreover, they use a variety of appropriate strategies to handle comprehension failures. These strategies are selected deliberately and deployed flexibly with regard to the particular text, purpose, and occasion (Baker & Brown, 1984; Paris, Wasik, & Turner, 1991; Pressley & Allington, 1999). Consequently, proficient readers are typified by both cognitive and metacognitive strategies, which are conscious, instantiated, and flexible plans readers apply and adapt to a variety of texts and tasks (Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989).

---

\* Based on: Van Keer, H., & Verhaeghe, J. P. Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. The present chapter is submitted for publication in the *Journal of Educational Psychology* and is currently under revision after review by the journal. Portions of the article were presented at the 9<sup>th</sup> EARLI-conference, Fribourg, Switzerland, August 28<sup>th</sup> - September 1<sup>st</sup> 2001.

Cognitive strategies can be defined as mental and behavioral activities, such as rereading, activating prior background knowledge, and adjusting reading speed, used to increase the likelihood of comprehending text (Van Den Broek & Kremer, 2000). On the other hand, metacognitive strategies can be specified as self-monitoring and self-regulating activities focusing on the products and the process of reading, helping readers' awareness of whether or not they comprehend what they are reading, and assisting readers' decision of what cognitive strategies to employ to aid comprehension as a function of text difficulty, situational constraints, and the reader's own cognitive abilities (Lories, Dardenne, & Yzerbyt, 1998; Van Den Broek & Kremer, 2000; Weisberg, 1988).

Unfortunately, there is no reason to believe that all elementary students spontaneously develop knowledge of the strategies, ability to select and apply appropriate strategies, and effective monitoring of strategies, just by exposure to varied reading material and by doing lots of reading (Hartman, 2001; Pressley & Allington, 1999). Reviews of research however reveal that monitoring and regulating skills, as well as the application of relevant cognitive strategies can be taught, leading to improved comprehension (Dole et al., 1991; Pressley, 2000; Pressley et al., 1989). In this respect, recent research underscores the significance of explicit reading comprehension instruction (Dole et al., 1991; Pearson & Fielding, 1991), including cognitive and metacognitive strategy instruction (Baumann, Seifert-Kessell, & Jones, 1992; Block, 1993; Dole, Brown, & Thraten, 1996), for "reading comprehension instruction takes the mystery out of the reading process, helping students assume control" (Raphael, 2000, p.76).

Contrary to the research results stressing the importance of explicit strategic reading comprehension instruction, there is evidence that little has changed since Durkin's (1978-1979) now classic observation research into reading comprehension instruction. The dominant instructional practice of teaching reading is still very traditional, characterized by questioning students about the content of a text after reading it, with little explicit attention to the strategic aspects of processing and comprehending text (Aarnoutse, 1995; Paris & Oka, 1986; Pressley, Wharton-McDonald, Hampston, & Echevarria, 1998; Weterings & Aarnoutse, 1986).

In addition to the significance of explicit reading strategies instruction, research demonstrates that children's development of reading competence in the elementary grades can be encouraged by providing them opportunities to enter in interaction about structured reading activities (Almasi, 1996; Fuchs, Fuchs, Mathes, & Simmons, 1997; Greenwood, Delquadri, & Hall, 1989; Mathes, & Fuchs, 1994; Johnson-Glenberg, 2000; Mathes, Torgesen, & Allor, 2001; Palincsar & Brown, 1984; Rosenshine & Meister, 1994; Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995). More particularly it is shown that through discussions, peer conferences, peer tutoring, and cooperative activities students implement, evaluate and modify strategy acquisition and use, and discuss how a strategy could be applied in situations other than the reading lessons (Klingner & Vaughn, 1996; Klingner, Vaughn, & Schumm, 1998; Palincsar & Brown, 1984). Moreover, discussions between peers provide opportunities for metacognitive exchanges and modeling (Palincsar, David, Winn, & Stevens, 1991). Two of the most prominent strategy-instruction programs encouraging social and task-oriented

interaction among peers in small reading groups are Reciprocal Teaching (Palincsar & Brown, 1984) and Transactional Strategies Instruction (Brown, Pressley, Van Meter, & Schuder, 1996; Pressley et al., 1992).

Notwithstanding the convincing research results, student-centered discussion with regard to reading comprehension is anything but common practice in most classrooms (Alvermann, 2000).

Taken into account the striking contrast between instructional practice and research evidence indicating the possibilities of explicit reading strategies instruction and peer-led interaction about texts, an intervention study was set up aiming at making reading comprehension instruction in primary schools in Flanders (Belgium) more effective. Two innovative cornerstones were included in the research, namely explicit instruction in reading comprehension strategies and opportunities to interact with peers about texts. More specifically, reading strategy instruction and peer tutoring were blended, for studies of paired peer tutoring in reading and thinking skills are relatively rare (Topping, 2001). With the exception of research on Peer-Assisted Learning Strategies (Fuchs, Fuchs, Mathes, et al., 1997), almost all of the research on tutoring students in reading has involved word-level oral reading or low-level comprehension activities rather than training in reading comprehension strategies (Pearson & Fielding, 1991).

Peer tutoring can be defined as "people from similar social groupings who are not professional teachers helping each other to learn, and learning themselves by teaching" (Topping, 1996, p. 322). This definition covers a whole series of practices that employ peers as one-on-one teachers to provide individualized instruction, practice, repetition and clarification of concepts (Topping, 1988; Utley & Mortweet, 1997). Peer tutoring is characterized by specific role taking: at any point someone has the job of tutor, while the other is in a role as tutee (Topping, 1996). Moreover, peer tutoring activities are structurally embedded in the curriculum and classroom organization and it is recommended that they are preceded by a special tutor preparation, for research demonstrates that peer tutoring is less effective when no attention is paid to a sound prior training of the tutors (Bentz & Fuchs, 1996; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997).

With regard to the composition of the dyads, two important categories of peer tutoring can be distinguished. Children can be paired with classmates. This variant is called same-age tutoring. The specific form of same-age tutoring in which the students alternate on a regular basis between the tutor and tutee role is called reciprocal same-age tutoring (Fantuzzo, King, & Heller, 1992). The second variant is called cross-age tutoring and refers to older students tutoring younger students.

Peer tutoring has been proven successful in a variety of curriculum areas and for different age groups. Research indicates positive effects on academic achievement, for both tutor and tutee (e.g., Cohen, Kulik, & Kulik, 1982; Fantuzzo, Davis, & Ginsburg, 1995; Fantuzzo, Polite, & Grayson, 1990; Fantuzzo et al., 1992; Greenwood, Carta, & Hall, 1988; Mathes et al., 2001; Simmons et al., 1995). As concerns the differential impact of same-age and cross-age tutoring, indirect reference material from the meta-analysis of Cohen et al. (1982) reveals that cross-age tutoring is more effective than

same-age tutoring, for the analysis shows larger effect sizes for both tutors' and tutees' achievement in cross-age tutoring programs. However, no statistically significant differences between the variants were found.

Positive effects of peer tutoring has also been found on tutors' and tutees' social and emotional functioning, more specifically with regard to improved self-efficacy perceptions, more positive self-concepts and social relationships, and better attitudes towards the curriculum areas treated in the tutoring sessions (e.g., Cohen et al., 1982; Fantuzzo et al., 1992; Fantuzzo et al., 1995; Greenwood et al., 1988; Mathes & Fuchs, 1994; Roswal et al., 1995). Especially the aspect of children's self-perceptions or self-efficacy judgments about one's ability to perform is an important construct with regard to reading comprehension, for attention to strategy instruction alone is not enough to produce maximum reading growth (Casteel, Isom, & Jordan, 2000). Affective factors result in deeper engagement with text, and this translates into superior reading achievement. In this respect, Henk and Melnick (1992, 1995) assert that self-efficacy judgments can affect an individual's overall orientation to the process of reading; influence choice of activities; affect continued involvement, amount of effort expended during reading, and the degree of persistence in pursuing text comprehension; and ultimately affect achievement.

The aim of the present intervention study was to create and evaluate complex sets of instructional interventions in real classrooms as tools for ameliorating second and fifth graders' reading comprehension achievement and self-efficacy perceptions towards reading. Contrary to prior research, merely focusing on either practicing reading strategies in small group work (e.g., Brown et al., 1996; Palincsar & Brown, 1984; Pressley et al., 1992) or peer tutoring for practicing lower level reading skills (Pearson & Fielding, 1991), the specific contribution of the present study is that it concentrates on peer tutoring as an instructional technique to practice the application of reading comprehension strategies. More specifically, the study enables an explicit comparison of the differential impact of practicing reading strategies in (a) teacher-led whole-class activities, (b) reciprocal same-age peer tutoring activities, or (c) cross-age peer tutoring activities, within the same study for two different age groups. Furthermore, the present study extends previous research by (a) sampling a relatively large number of participants from 44 classrooms in 25 schools throughout Flanders; (b) supporting teachers to implement the innovations in their regular classroom context with the participation of all students in the course of an entire school year; (c) including long-term maintenance assessments; (d) using standardized reading comprehension tests not linked directly to the treatment; and (e) applying multilevel modeling to take the hierarchical nesting of students in classes into account.

Based on a review of research literature, the major hypotheses of the study can be formulated as follows:

1. Explicit reading strategies instruction, followed by practice in teacher-led whole-class activities or peer tutoring activities, enhances second and fifth graders' reading comprehension achievement more than traditional reading comprehension instruction.



2. Cross-age and reciprocal same-age peer tutoring activities during which reading strategies are practiced, generate greater positive changes in second and fifth graders' reading comprehension achievement than the more traditional teacher-led practice of these strategies during whole-class activities.
3. The improvement in reading comprehension is more obvious for second graders functioning as tutees in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities.
4. The improvement in reading comprehension is more obvious for fifth graders functioning as tutors in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities.
5. Cross-age and reciprocal same-age peer tutoring activities improve second and fifth graders' self-efficacy perceptions towards reading more than traditional teacher-led instructional techniques.
6. The improvement in self-efficacy perceptions towards reading is more obvious for second graders functioning as tutees in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities.
7. The improvement in self-efficacy perceptions towards reading is more obvious for fifth graders functioning as tutors in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities.

In addition to the seven hypotheses, we explored to what degree the intervention effects on reading comprehension achievement mediate changes in positive or negative self-efficacy perceptions with regard to reading and associated success or failure attributions. The possible inverse mediating effects of the students' on-line self-efficacy thoughts and associated causal attributions on their growth in reading comprehension were also explored.

## Method

### *Design*

A pretest posttest and retention test control group design was used in the study. To ensure the ecological validity of the interventions, complete naturally composed classes were included, so the design can be referred to as quasi-experimental. This setting provides a more stringent test of the successful implementation of the interventions than studies in tightly controlled laboratory settings, of which research results can not simply be transferred to the context of real-life classrooms. To the extent that the results show that the interventions are effective in the quasi-experimental context, they will strengthen their applicability and effectiveness in natural classroom settings.

Participating classes were assigned to four research conditions. In the strategies-only condition (STRAT) the experimental intervention included explicit instruction in reading strategies, followed by practicing the strategies in teacher-led whole-class

settings. The experimental same-age (STRAT+SA) and cross-age (STRAT+CA) peer tutoring condition included identical explicit instruction in the same reading strategies, but the strategies were practiced in respectively reciprocal same-age and cross-age dyads. Peer tutoring activities were organized class-wide, so all students in a class were paired and worked simultaneously. Finally there was a control group, characterized by traditional reading comprehension instruction without explicit instruction or peer tutoring. The effects of peer tutoring activities independent of teacher-directed explicit reading strategies instruction were not investigated, because in traditional reading comprehension instruction no explicit instruction is provided, so there is no content for tutoring.

### *Participants*

In total 444 second-grade and 454 fifth-grade students from 44 classes in 25 different schools throughout Flanders participated in the study. Except for one inner-city school (one second- and fifth-grade class in the STRAT condition) with mainly a low SES and ethnic minority population, all schools had a predominantly white, Flemish population. Although principally having a population that is mixed with regard to SES, the majority of the children are from middle-class families. Except for one second-grade class including only girls, there was approximately an even division of gender: on average 53% ( $SD = 16.54$ ) and 48% ( $SD = 18.55$ ) of the students were boys in respectively the second- and fifth-grade classes. At the beginning of the school year second graders' age ranged from 6 to 9 years, with an average of 7 years and 4 months, while fifth graders' age ranged from 9 to 12, with an average of 10 years and 5 months. The majority of the students (402 in second grade and 422 in fifth grade) were native Dutch speakers, which is the medium of instruction in Flanders. Classes are to be considered as academically heterogeneous. Class size ranged from 15 to 28 with an average of almost 21 ( $SD = 3.50$ ) in second grade and from 10 to 30 in fifth grade with on average almost 22 ( $SD = 5.00$ ) students per class, which is representative for the Flemish situation.

Second-grade teachers were on average 33 years old and had 11 years of teaching experience. Fifth-grade teachers were somewhat older, with a mean age of 42 and almost 20 years of teaching experience. Four second-grade and five fifth-grade teachers were men. None of the teachers had experience in the experimental interventions.

Participating teachers were selected from a group of about 100 second- and fifth-grade teachers, all willing to take part in a long-term research study. The selection was based on the geographical distribution of the schools throughout the whole of Flanders on the one hand, and on the possibility to match teachers and classes as closely as possible with regard to teachers' teaching experience, class size, students' age, gender distribution, and dominating mother tongue on the other hand. Teachers were randomly assigned to the experimental STRAT or tutoring conditions. Within the peer tutoring conditions they were allowed to opt in favor of the STRAT+SA or STRAT+CA condition according to their own preference and the readiness of a

second- or fifth-grade colleague to collaborate in the STRAT+CA activities. Control group classes were selected to match the teachers and class groups in the experimental conditions.

Table 2.1 shows the number of classes and students in each condition for both second and fifth grade. The fact that twice the number of classes were assigned to the STRAT condition as compared to the STRAT+SA and STRAT+CA condition had to do with the continuation of the present study, intended to replicate the implementation of both tutoring conditions with a larger number of classes.

Table 2.1  
*Number of Participating Classes and Students*

Condition	Grade			
	2 <sup>nd</sup>		5 <sup>th</sup>	
	Classes	Students	Classes	Students
STRAT+SA	5	91	4	101
STRAT+CA	3	66	4	69
STRAT	8	163	8	177
Control group	6	124	6	107
Total	22	444	22	454

### *Measurement Instruments*

The instruments used in the study were standardized tests measuring the reading comprehension achievement and the decoding fluency of the students, as well as questionnaires with respect to reading attitude, perceived competence, and preoccupation with attributions and self-efficacy perceptions towards reading.

### *Reading Comprehension Tests*

In second grade reading comprehension achievement was measured using Dutch standardized test batteries, which are well-established and widely used instruments for the assessment of students' reading comprehension. At each measurement occasion a different test, increasing in level of difficulty, was used to collect the data. For the pretest a test from the test battery "Lezen met begrip 1" (Reading with comprehension) (Verhoeven, 1993) was utilized. This test contains six short stories, each followed by five multiple-choice questions respectively asking for the meaning of a word, the meaning of a sentence, the referral relation between words, the connection between sentences, and the theme of a text. The scores are determined by summing up the correct answers. For the post- and retention test tests from the battery "Toetsen Begrijpend Lezen" (Reading comprehension tests) (Staphorsius & Krom, 1996) were applied. These tests consist of respectively four and three different stories followed by in total 25 multiple-choice questions. More specifically, two types of questions can be distinguished, that is questions concerning the content of a text (demanding a clear understanding of the meaning of words and sentences, the referral relation between words, the connection between sentences, and the theme of the text) and questions concerning the communication between the author and the reader of the text (e.g.,

objective of the author, intended target group, the author's attitude towards the matter he raises), both requiring the integration of information on different textual levels (words, sentences, successive sentences, and the text in its entirety). After discussing an example, students completed the tests individually. To examine the internal consistency of the tests Cronbach's  $\alpha$ -coefficients were calculated for the three measurements occasions, yielding high reliability scores of .90 ( $n = 432$ ) for the pretest, .84 ( $n = 422$ ) for the posttest, and .83 ( $n = 385$ ) for the retention test.

In fifth grade reading comprehension achievement was measured with the standardized test battery "Toetsen Begrijpend Lezen" (Reading comprehension tests) (Staphorsius & Krom, 1996). At each measurement occasion a different test, increasing in level of difficulty, was used to collect the data. More specifically, the tests designed for fourth, fifth and sixth grade were used to collect the data. In common with second graders' post- and retention test, two types of questions, both requiring the integration of information on different textual levels, can be distinguished, that is questions concerning the content of a text and concerning the communication between the author and the reader. After discussing an example, students completed the tests individually. The tests consist of three modules of 25 multiple-choice questions each. All students took the first module of the test. Afterwards they completed the second more easy or third more difficult module, depending on their first results. The scores are determined by summing up the correct answers. To compare the scores of the students who finished the more easy or more difficult part of the test, the sum scores were transposed into an IRT-modeled global achievement score following the test guidelines. To verify the reliability of the three modules of the pre-, post-, and retention test, Chronbach's  $\alpha$ -coefficients were computed. Table 2.2 indicates that all reading comprehension measures are reliable.

Table 2.2

*Chronbach's  $\alpha$ -Coefficients of the Fifth-Grade Reading Comprehension Tests*

Test module	Measurement occasion <sup>a</sup>		
	Pretest	Posttest	Retention test
1	.81 ( $n = 468$ )	.72 ( $n = 442$ )	.76 ( $n = 403$ )
2	.76 ( $n = 167$ )	.76 ( $n = 256$ )	.79 ( $n = 362$ )
3	.66 ( $n = 271$ )	.74 ( $n = 175$ )	.77 ( $n = 41$ )

<sup>a</sup> At each measurement occasion a different test, increasing in level of difficulty, was used.

### *Decoding Fluency Test*

The "Eén minuut test" (One minute test) (Brus, 1969) was administered individually to all second graders at each measurement occasion to assess their decoding fluency, which is a combination of accuracy and decoding speed (Chard, Simmons, & Kameenui, 1998). In this standardized test students read unrelated words with an increasing level of difficulty during exactly one minute. The score is determined by counting the number of words read correctly.

### *Reading Attitude Scale*

In both second and fifth grade, children completed a Dutch Reading Attitude Scale (Aarnoutse, 1996) at pre- and posttest. Fifth graders read and completed the questionnaire individually. In second grade all 27 items of the scale were read out loud by the first author and completed individually by the students. Reading attitude can be unfolded in three components, that is children's knowledge and experience with reading, their positive or negative appreciation of reading and reading material, and their tendency to read. The Reading Attitude Scale refers to these different elements, but focuses chiefly on the affective aspect (Aarnoutse, 1996). The score on the scale is measured as the number of positive judgments on the 27 questions. To examine the internal consistency of the scale Cronbach's  $\alpha$ -coefficients were calculated for both second and fifth graders, yielding high reliability scores of respectively .84 ( $n = 334$ ) and .94 ( $n = 368$ ) for the pretest, and .92 ( $n = 353$ ) and .92 ( $n = 395$ ) for the posttest.

### *Perceived Competence Scale*

In fifth grade the "Competentiebelevingsschaal voor Kinderen (CBSK)" (Veerman, Straathof, Treffers, Van den Bergh, & ten Brink, 1997), a Dutch version of the Self-Perception Profile for Children (Harter, 1985), was administered at pre- and posttest to measure children's self-perceptions in relation to specific domains of one's life (scholastic competence, athletic competence, social acceptance, physical appearance, and behavioral conduct) and a separate facet of global self-worth. The CBSK is a 36-item self-report questionnaire divided into the six subscales mentioned above. For example, to assess self-perceived scholastic competence, children were asked to first choose which of the following sentences sounded more like them: "Some children feel that they are very good at their school work" or "Other children worry about whether they can do the school work assigned to them." After the children had decided which sentence sounded more like themselves, they were then asked to judge whether the statement was "really true" or "sort of true" for them. All questions followed the same format. To verify the reliability of the different scales Chronbach's  $\alpha$ -coefficients were computed. As can be seen in Table 2.3 the reliability of the measures is acceptable. Because the questionnaire was not suitable for second graders, this instrument was not used with this age group.

Table 2.3  
*Chronbach's  $\alpha$ -Coefficients of the Fifth-Grade CBSK-scales*

CBSK-scale	Measurement occasion	
	Pretest	Posttest
Scholastic competence	.76 ( $n = 428$ )	.78 ( $n = 426$ )
Social acceptance	.71 ( $n = 432$ )	.80 ( $n = 421$ )
Athletic competence	.77 ( $n = 430$ )	.77 ( $n = 419$ )
Physical appearance	.85 ( $n = 437$ )	.88 ( $n = 424$ )
Behavioral conduct	.71 ( $n = 431$ )	.74 ( $n = 418$ )
Global self-worth	.76 ( $n = 434$ )	.81 ( $n = 428$ )

### *Questionnaire on Self-Efficacy Perceptions and Related Causal Attributions*

Within the framework of the present study, a questionnaire was developed to measure how often students are preoccupied with positive or negative thoughts with regard to their own reading ability or related causal attributions. Inspired by the work of Ames (1984), children were asked to report how often such thoughts crossed their mind either before, while, or after reading by ticking one of the boxes ("never", "almost never", "sometimes", "very often") following statements as: "I think: I comprehended well because it was an easy text.", "I think: I did not comprehend well because there was no one to help me.", "I think: I am not good at reading". Factor analysis revealed that success attributions and positive thoughts about one's own reading competence on the one hand and failure attributions and negative self-efficacy perceptions with regard to reading on the other hand are very closely related. This result is in line with the findings of Marsh (1984) and Marsh, Cairns, Relich, Barnes, and Debus (1984) that self-attributions can be seen as expressions or indicators of one's self-concept or self-efficacy perceptions. Therefore, two scales were constructed. It is important to mention that a low score on the scale with regard to positive thoughts relating to self-efficacy does not necessarily mean that a student has a low self-esteem with regard to reading proficiency. It only reveals that the student is not preoccupied with thoughts about his reading proficiency or success. On the other hand, a high score on the scale with regard to negative thoughts relating to self-efficacy in reading clearly indicates the presence of a low self-esteem with regard to reading.

The questionnaire was administered at each measurement occasion to both second and fifth graders. Fifth graders read and completed the questionnaire individually. In second grade all items were read out loud by the first author and judged individually by the students. Table 2.4 presents the results of the internal consistency analyses, revealing that the combination of the thoughts relating to success and failure attributions with respectively the positive and negative self-efficacy perceptions towards reading are on an acceptable level of reliability.

Table 2.4

*Chronbach's  $\alpha$ -Coefficients of the Questionnaire concerning Preoccupation with Attributions and Self-Efficacy Perceptions*

Questionnaire scale	Measurement occasion			
	Pretest		Posttest	
	2 <sup>nd</sup> grade	5 <sup>th</sup> grade	2 <sup>nd</sup> grade	5 <sup>th</sup> grade
Success attributions and positive self-efficacy perceptions	.63 (n = 419)	.69 (n = 441)	.75 (n = 402)	.71 (n = 426)
Failure attributions and negative self-efficacy perceptions	.77 (n = 367)	.83 (n = 408)	.84 (n = 368)	.81 (n = 393)

### *Interventions*

#### *Support to the Teachers*

Because the interventions in the three experimental conditions (STRAT, STRAT+SA, and STRAT+CA) were not implemented by the researchers but by the regular classroom teachers, the first author trained teachers in all three experimental

conditions to ensure that they were implementing the experimental interventions with fidelity. To that end, teachers in the experimental conditions were provided with an elaborated manual including all materials necessary to conduct the innovation, namely (a) a description of the objectives and the organization of the interventions, (b) lesson scenarios describing the objectives, the necessary materials, and the successive phases of each lesson, and (c) supplementary student materials, such as assignment cards and reading texts. Teachers were not required to develop additional materials. Moreover, teachers were provided with in-service training, technical assistance and coaching by the first author. Prior to the interventions in the classrooms, the underlying theoretical background of the innovations was clarified, an overview and the organization of the interventions was outlined, and the provided manual and additional materials were fully discussed during local meetings in the schools. Subsequently, after the start of the interventions, monthly discussions took place to exchange experiences and ideas, and to overcome practical or implementation difficulties. Two structured interviews and observations were carried out to document the fidelity of the interventions.

#### *STRAT+SA and STRAT+CA Condition*

With regard to the experimental interventions both the STRAT+SA and STRAT+CA conditions are characterized by three elements, that is explicit instruction in reading strategies, a sound tutor preparation, and practice of the application of reading strategies in weekly peer tutoring sessions.

*Explicit instruction in reading strategies.* Concerning the explicit instruction in reading strategies teachers were provided with elaborated lesson scenarios and materials for teaching students to apply six essential strategies: (a) activating prior background knowledge and linking it to the text, (b) predictive reading and checking story outcomes, (c) distinguishing main issues from side-issues, (d) monitoring and regulating the understanding of words and expressions, (e) monitoring and regulating comprehension by tracing the ideas expressed in difficult and not understood sentences or passages, and (f) classifying types of text and adjusting reading behavior to it. The compilation of these reading strategies was based on feasibility and relevance for both second and fifth graders, and inspired by contemporary reading research and recurrent strategies in programs focusing on explicit strategy instruction (e.g., Brown et al., 1996; De Corte, Verschaffel, & Van De Ven, 2001; Fuchs, Fuchs, Mathes, et al., 1997; Fukkink, Van der Linden, Vosse, & Vaessen, 1997; Klingner & Vaughn, 1996; Palincsar & Brown, 1984). Activating prior knowledge involves that students evoke what they already know about the topic of the text before they start reading and is generally considered as crucial for text comprehension (Palincsar & Brown, 1984; Paris & Oka, 1986; Pressley et al., 1992; Walraven, 1995). Within the scope of the intervention, students were more particularly instructed to infer the topic of the text on the basis of the title and accompanying illustrations, ask themselves what they already knew about that topic, and note this down in a number of key words. As regards the second reading strategy, students were taught to make predictions about what will be read on the next pages of a text and to compare these predictions to the actual story outcomes. In this way, students are forced to look ahead while reading and to verify expectations afterwards. Distinguishing main issues from side-issues targets the skills

of main idea identification and summarizing. In the present study, students more specifically learned to ask and answer who-, what-, where-, when-, and why-questions in reference to the text and to restate the main ideas of successive paragraphs of the text or of the full text. Monitoring and regulating the understanding of difficult words and expressions is essential for understanding texts as well. Students were taught to identify not understood words or expressions and to discover the meaning by looking for a definition, a synonym, or a description in the text, by deriving clarification from the context, by referring to a dictionary or computer, or by enlisting someone's help. On the basis of the fifth strategy, students were encouraged to monitor and check their comprehension and to regulate understanding of difficult sentences or passages by rereading, adjusting reading speed, or tracing the meaning of unfamiliar words or expressions. Finally, students were alerted to the fact that different types of texts - each with their own characteristics - can be distinguished. They were instructed in classifying types of text and adjusting reading behavior to it.

Apart from a selection of text excerpts to practice the strategies, the materials provided to the students also included assignment cards, offering structure and visual support during the reading process by displaying step by step how to employ the reading strategies. The assignment cards were used by the teachers when first explaining the strategies and remained later as students' reference.

In designing the lesson scenarios components of transactional strategies instruction (Brown et al, 1996; Pressley et al., 1992) and reciprocal teaching (Palincsar & Brown, 1984) were taken into account. During a whole-class instructional presentation phase of each strategy much attention was paid to extensive and direct teacher explanations and modeling of strategic reasoning. The teacher explicitly explained and modeled by using the think-aloud methodology why, how, and when a specific strategy can be helpful in enhancing comprehension. In this way, declarative (What is the nature of the strategy?), procedural (How to deploy it?) and conditional (When to use it?) knowledge (Paris, Lipson, & Wixson, 1983) of the strategies was imparted to the students. In addition, a gradual transfer from external regulation by the teacher to self-regulation of strategy use by the students was taken into account. During a phase of practice and coaching using multiple examples the teacher put the strategies into practice together with the students. This phase was characterized by student assignments, systematic and explicit scaffolding and coaching by the teacher to engage students in applying and reflecting upon the strategies, and subsequent whole-class discussions. In a last phase the assignment cards were applied more independently to systematize the use of the strategies. In the STRAT+CA and STRAT+SA conditions the independent practice took place in respectively cross-age and reciprocal same-age peer tutoring dyads.

*Tutor preparation.* Apart from the explicit instruction in reading strategies, both tutoring conditions were characterized by a preparation of the tutors. A series of lessons and materials were developed to train students in becoming a good tutor, based on related research and tutoring programs (Bentz & Fuchs, 1996; Fukkink et al., 1997). The lesson scenarios and student materials were included in STRAT+CA and STRAT+SA teachers' manual and comprised instructions for role play and worksheets, as well as an administration card on which the tutors were expected to write down



what texts they read with their tutees and which assignments had been practiced during the sessions.

The preparatory lessons were scheduled at the beginning of the intervention and required seven 50-minute sessions. Tutors more particularly got acquainted with their tasks and responsibilities and they learned how to show interest, how to initiate and finish a session, how to give corrective feedback and reinforcement, how to provide praise and how to offer explanations and assistance.

*Peer tutoring sessions.* Finally, both tutoring conditions were typified by the innovative cornerstone of the peer tutoring activities. In the STRAT+CA condition fifth graders were paired with younger students of the second grade. Children were assigned to dyads by their classroom teachers. Besides personality of the children, dyad composition was especially based on the children's reading ability so that poor and good readers of fifth grade were respectively paired with poor and good readers of second grade. In the STRAT+SA condition second and fifth graders were paired with classmates. Teachers assigned students to academically heterogeneous and socially compatible pairs. In the case of an uneven number of students in the class, one group of three children was put together or a student from a parallel class was added to the intervention project. Important in the STRAT+SA condition is that students in the same-age dyads alternated regularly between tutee and tutor roles, so that both served as tutor for an equal amount of time. For this reason all second and fifth graders in the reciprocal STRAT+SA condition were prepared to take on the tutor role. In the STRAT+CA condition only fifth graders attended the preparatory tutor training. In principle dyads remained together for the duration of the school year. However, the couples were changed in the case of socially incompatible pairs.

During the peer tutoring activities an important role was reserved for the teachers. They were present and circulated around the classroom at all times to observe and coach the reading couples. Apart from the preparatory tutor training and the introductory lessons on reading strategies, peer tutoring sessions were organized once or twice a week on the average and lasted from 25 to 50 minutes, depending on the task and scheduling.

With regard to the succession of the introductory lessons and the independent practice in peer tutoring sessions, a "sandwich model" was applied. This implies that each lesson in which a new reading strategy was introduced, was followed by at least one peer tutoring session in which that strategy was practiced when reading one of the selected text excerpts from the manual. Thereafter a number of tutoring sessions followed in which the strategy was practiced when reading books or texts the students could choose themselves from the school or class library. Taken into account prior research results stressing that a combination of dyadic peer interaction and structured academic activity do more to enhance cognitive gain than either of the two dimensions separately (Cohen et al., 1982; Fantuzzo, Riggio, Connelly, & Dimeff, 1989; Lambiotte et al., 1987), the aforementioned strategy assignments cards were used as a vehicle for structuring the peer tutoring interaction about the application of the reading strategies. At the end of the school year students were not only expected to make their own choice of reading books, but also to make their own choice of relevant reading strategies and according assignment cards to work on. Prior to each peer tutoring

session a short briefing was given to the students pointing out the main ideas discussed in the previous introductory lessons. Moreover, all peer tutoring sessions were followed by a short reflection in which students' experiences and teachers' observations were discussed. During the in-service training, teachers were provided with guidelines for putting the reading dyads together, organizing the peer tutoring sessions, and coaching the reading dyads.

### *STRAT Condition*

The STRAT condition is typified by identical explicit instruction in the same six reading strategies as the STRAT+SA and STRAT+CA conditions. Teachers were not only provided with a manual including the same lesson scenarios and the same assignment cards and text excerpts, they also build in the gradual transfer of responsibility from teacher to students. The practice to systematize the use of the strategies did however not occur in peer tutoring reading dyads, but was characterized by individual seatwork, teacher-led discussions, and no highly interactive peer-mediated instructional techniques. To alternate the introductory strategy lessons and the practice sessions, a sandwich model was used as well: each introduced reading strategy was followed by at least one training session applying one of the text excerpts in the manual and a number of training sessions on teacher-selected books or texts.

### *Control Condition*

The control group teachers conducted reading instruction in their typical fashion, that is without explicit instruction in reading strategies or peer tutoring. Interviews with the teachers of these classes more specifically revealed that the lessons typically involved teacher-led whole-class activities, including asking comprehension-check questions after reading a text, teacher evaluation of students' answers, and presentation of the correct answers. Control group teachers were told that the purpose of the study was to examine primary school students' development in the field of reading comprehension achievement and they were not informed that they were part of a control condition.

### *Time Spent on Reading Comprehension Instruction*

The interventions in the experimental conditions were not meant as an additional program on top of teachers' traditional reading comprehension classes. Experimental teachers were coached to substitute their traditional way of teaching reading comprehension with one of the experimental programs and they were encouraged to implement the treatments during time normally allocated for reading instruction. Therefore, no differences between the four conditions were intended with regard to the total amount of time spent on reading comprehension instruction and practice.

The structured interviews with the experimental teachers yielded information on the total amount of time they spent on reading comprehension instruction on a weekly and yearly basis. The same information was obtained from all control group teachers. One-way analyses of variance on these data revealed that the different conditions in both the second ( $F = 1.188$ ,  $df = 3$ ,  $p = .342$ ) and fifth grade ( $F = 0.348$ ,  $df = 3$ ,  $p = .791$ ) allocated commensurable amounts of time to reading comprehension instruction. Post hoc analyses neither produced mutual significant differences between the conditions.

### *Procedure*

The research ran from September 1999 to December 2000. The implementation of the experimental interventions was spread out over the entire school year 1999-2000 and was conducted with all students during regularly scheduled reading instruction.

Data were collected by the first author within the regular classroom context and during regularly scheduled class sessions. All four conditions were measured at three points in time, that is a pretest in October (second and fifth grade) before the initiation of the intervention program, a posttest in the period of May and June (second and fifth grade) after the completion of the experimental intervention, and a retention test in December (third and sixth grade). It is to be stressed that, in accordance with the planning of the research, none of the third- and sixth-grade teachers pursued the experimental intervention. Consequently, between post- and retention test (i.e., in the first term of third and sixth grade) all participants got traditional reading instruction and no experimental intervention was organized.

### *Data Analysis*

In this study students are nested within a smaller number of classes. Therefore we can argue that the problem under investigation has a clear multilevel structure. A consequence of this hierarchical structure is that the observations of individual students are generally not completely independent because of the common history and experiences children share by belonging to the same class (Hox, 1994). For this reason the application of hierarchical or multilevel models is recommended, because these techniques - in contrast to the traditional ordinary least squares regression analysis - take the interdependency explicitly in account (Bryk & Raudenbush, 1992; Hox & Kreft, 1994).

In the present article two levels were distinguished: students (level 1) are clustered within classes (level 2). Because the number of class groups within schools ranged from one to two for each grade, the school level was not taken into account. Separate analysis were done for second and fifth graders. To facilitate the interpretation of the estimates, standardized scores were calculated and included in the models. The parameters of the models were estimated using the iterative generalized least squares (IGLS) estimation procedure of the software MlwiN (Rasbash et al., 1999).

## **Results**

### *Effects on Reading Comprehension*

Tables 2.5 to 2.8 present the results concerning reading comprehension achievement for second and fifth graders' post- and retention test respectively. With a view to build the most appropriate models and to test the research hypotheses, each model was built up from a null model to a model including relevant explanatory variables, whereupon the effects attributable to the experimental interventions were explored.

The first step in the analyses was to examine the results of an unconditional two-level null model, with only an intercept term included (Model 0). This model serves as a baseline with which to compare subsequent more complex models and is unconditional because the variance components are not predicted by any variables. Moreover, the null model permits partitioning the total variance into within-class and between-class components. The random part of Model 0 in Table 2.5 to 2.8 provides justification for applying multilevel models, for the variances at the class and student level of second and fifth graders' post- and retention test scores are all significantly different from zero. Respectively 11% and 15% of the total variance in second graders' post- and retention test scores is connected to differences between classes. For fifth graders respectively 13% and 20% of the total variance in post- and retention test scores can be attributed to differences between classes. As standardized scores were included in the analysis, the null models' intercepts - which represent the overall mean score of all students in all classes - are not significantly different from zero.

The second step in the construction of the models concerned the inclusion of conceivably relevant explanatory variables, such as pretest measures and students' background characteristics. The class-level total amount of time spent on reading comprehension instruction was also introduced as a potential predictor. However, the inclusion of this variable revealed no significant effect in none of the analyses. Since parsimonious models are preferred, only significant predictors ameliorating the model were retained.

As Model 1 in Table 2.5 to 2.8 reveals, students' pretest reading comprehension achievement is an important significant predictor of both second and fifth graders' post- and retention test scores, with effect sizes ranging from 0.59 to 0.76 standard deviation. When adding second graders' pretest scores to the fixed part of the model, the random part estimates reveal that the great majority of the post- and retention test differences between classes (respectively 86% and 71%) and between students (respectively 43% and 33%) is accounted for by pretest differences. The same is true for fifth graders. Adding fifth-grade students' pretest scores to the models explains an important part of the post- and retention differences between classes (respectively 61% and 44%) and between students (respectively 61% and 48%).

As it is possible that the level 1 and level 2 variances in second and fifth graders' post- and retention test scores differ according to the students' pretest measures, we allowed the parameter estimates of this predictor to vary randomly across all classes and students. Only for fifth graders' posttest scores the pretest parameter estimate showed complex variance at level 1. More specifically, the variance in posttest scores between students within a class can be considered as a linear function of the pretest scores and increases as pretest scores go up. This implies that the differences in posttest reading comprehension achievement between fifth graders within a class expand as students performed better on the pretest.

Apart from students' pretest reading comprehension achievement, a number of other variables appeared to be significant predictors of second and fifth graders' post- and retention test scores. As concerns second graders' post- and retention test scores, the fixed part parameter estimates of Model 2 in Table 2.5 and 2.6 indicate that higher decoding fluency scores at pretest go together with higher reading comprehension

scores at post- and retention test, whereas the number of years children are behind at school and the pretest measure of preoccupation with failure attributions towards reading are inversely proportional to the post- and retention test comprehension scores. As regards the impact of second graders' gender, we can conclude that having controlled for relevant pretest measures, boys still perform significantly lower at both post- and retention test. Since these relevant pretest measures include previous reading comprehension achievement, it can be concluded that with regard to reading comprehension seven-year-old boys tend to grow significantly slower than girls. As can be seen in Model 2 in Table 2.6, an additional significant background characteristic was included in second graders' retention test model, revealing that non-native speakers perform and grow significantly less than native Dutch-speaking children.

With regard to fifth graders' post- and retention test scores, the fixed part estimates of Model 2 in Table 2.7 and 2.8 indicate that, apart from pretest reading comprehension achievement, higher pretest measures of personal attitude towards reading and perceived scholastic competence go together with significantly higher post- and retention test scores. With regard to the explanatory variables for the posttest scores, it has to be mentioned that the influence of an additional variable, namely the average pretest reading attitude per class, is not statistically significant in Model 2. However, after adding the dummies representing the three experimental conditions (Model 3) this variable does become significant. Therefore, the variable was added to the model.

Similar to the analyses for Model 1, the assumption of complex variances at level 1 or 2 was examined for the estimates of all the additional significant predictors of second and fifth graders' post- and retention test scores. Only for the estimates of the effects of pretest decoding fluency, these analyses indicated complex level 1 variance in second graders' post and retention test scores. More specifically, the random part estimates of Model 2 in Table 2.5 and 2.6 reveal that the variance in post- and retention test reading comprehension achievement between students within classes decreases at a constant rate as pretest scores increase. This implies that the within class differences between second graders for post- and retention test reading comprehension achievement were less large for students performing better on the pretest decoding fluency test.

To test the research hypotheses, the third and final step in the analyses was to examine the effects attributable to the experimental interventions by adding the categorical variable "condition" to the model, controlling for the relevant predictors mentioned above (Model 3). To represent the four research conditions three dummy-variables were used, with the control group as reference category.

The parameter estimates of Model 3 in Table 2.5 reveal significant effects for the STRAT and STRAT+CA condition, entailing that by the end of the school year second-grade students in these conditions had made significantly more progress in reading comprehension compared to children in the control group, with effect sizes of respectively 0.23 and 0.22 standard deviation. No significant effects were found for the STRAT+SA condition, and pairwise comparisons neither revealed significant differences between the three experimental conditions. Furthermore, no interaction effects occurred on the posttest between condition and gender, nor between condition and initial performance level. These results imply that the significant positive effects

of the STRAT and STRAT+CA conditions are equally strong for boys as for girls and for initially low scoring children as for high achievers.

Table 2.5

*Model Estimates for the Two-Level Analyses of the Second-Grade Posttest Reading Comprehension Achievement Scores*

Parameter	Model			
	0	1	2	3
<i>Fixed</i>				
Intercept	-0.00 (0.09)	0.01 (0.05)	0.10 (0.06)	-0.04 (0.08)
Pretest comprehension		0.68 (0.04)***	0.58 (0.05)***	0.57 (0.05)***
Pretest reading fluency			0.09 (0.05)*	0.10 (0.05)*
Gender (boy)			-0.16 (0.07)*	-0.16 (0.07)*
Years behind at school			-0.27 (0.09)**	-0.30 (0.09)**
Pretest failure attributions			-0.11 (0.04)**	-0.11 (0.04)**
STRAT+SA				0.14 (0.11)
STRAT+CA				0.22 (0.11)*
STRAT				0.23 (0.08)**
<i>Random</i>				
Level 2				
$\sigma^2_{\mu 0}$	0.11 (0.05)*	0.02 (0.01)	0.01 (0.01)	0.00 (0.00)
Level 1				
$\sigma^2_{e0}$	0.88 (0.06)***	0.50 (0.04)***	0.46 (0.04)***	0.46 (0.03)***
$\sigma_{e0\text{fluency}}$			-0.04 (0.02)**	-0.04 (0.02)*
<i>Deviance</i>	1175.30	895.70	767.27	759.71

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

Table 2.6

*Model Estimates for the Two-Level Analyses of the Second-Grade Retention Test Reading Comprehension Achievement Scores*

Parameter	Model			
	0	1	2	3
<i>Fixed</i>				
Intercept	-0.02 (0.10)	-0.05 (0.06)	0.15 (0.08)	0.12 (0.11)
Pretest comprehension		0.59 (0.04)***	0.47 (0.06)***	0.47 (0.06)***
Pretest reading fluency			0.14 (0.05)**	0.14 (0.05)***
Gender (boy)			-0.21 (0.08)*	-0.19 (0.08)*
Years behind at school			-0.23 (0.21)*	-0.24 (0.11)*
Non-native speaker			-0.36 (0.18)*	-0.36 (0.18)*
Pretest failure attributions			-0.09 (0.04)*	-0.09 (0.04)*
STRAT+SA				0.24 (0.17)
STRAT+CA				-0.06 (0.16)
STRAT				-0.00 (0.13)
<i>Random</i>				
Level 2				
$\sigma^2_{\mu 0}$	0.15 (0.06)**	0.04 (0.02)	0.03 (0.02)	0.02 (0.02)
Level 1				
$\sigma^2_{e0}$	0.85 (0.06)***	0.57 (0.04)***	0.54 (0.04)***	0.54 (0.04)***
$\sigma_{e0\text{fluency}}$			-0.06 (0.02)***	-0.06 (0.02)***
<i>Deviance</i>	1062.20	872.18	746.54	743.73

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

Contrary to the posttest results, on the retention test no significant differences were found for second grade between any of the experimental conditions and the control group (see Table 2.6, Model 3). It is clear that the positive effects of explicit instruction in reading strategies and practice in cross-age peer tutoring dyads did not last once the experimental intervention was finished.

With regard to fifth graders' posttest scores, Model 3 in Table 2.7 demonstrates significant positive effects for all three experimental conditions with effect sizes ranging from 0.32 to 0.39 standard deviation, indicating that the growth in reading comprehension produced by the experimental interventions was considerably higher than the growth observed in the control group. Pairwise comparisons revealed no mutual significant differences between the experimental conditions. Furthermore, no interaction effects occurred on the posttest between condition and students' initial performance level. This implies that the significant positive effect of the experimental conditions is equally strong for initially low scoring children as compared to high achievers.

Table 2.7

*Model Estimates for the Two-Level Analyses of the Fifth-Grade Posttest Reading Comprehension Achievement Scores*

Parameter	Model			
	0	1	2	3
<i>Fixed</i>				
Intercept	-0.04 (0.09)	0.01 (0.06)	0.01 (0.05)	-0.25 (0.09)**
Pretest comprehension		0.76 (0.03)***	0.70 (0.03)***	0.70 (0.03)***
Pretest individual attitude			0.06 (0.03)*	0.06 (0.03)*
Pretest mean class attitude			0.25 (0.19)	0.37 (0.16)*
Pretest perceived scholastic competence			0.08 (0.03)*	0.07 (0.03)*
STRAT+SA				0.39 (0.14)**
STRAT+CA				0.32 (0.15)*
STRAT				0.33 (0.12)**
<i>Random</i>				
Level 2				
$\sigma^2_{\mu 0}$	0.13 (0.05)*	0.05 (0.02)*	0.05 (0.02)*	0.03 (0.01)*
Level 1				
$\sigma^2_{\epsilon 0}$	0.87 (0.06)***	0.34 (0.03)***	0.33 (0.02)***	0.33 (0.02)***
$\sigma^2_{\epsilon 0 \text{comprehension}}$		0.03 (0.01)*	0.03 (0.01)**	0.03 (0.01)**
<i>Deviance</i>	1191.90	759.03	723.82	714.69

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

The most striking result was found with regard to fifth graders' retention test scores (see Table 2.8, Model 3), for which a quite large significant effect of the STRAT+CA condition was found with an effect size of 0.60 standard deviation, indicating that with respect to reading comprehension fifth graders who acted as tutors for second-grade students continued growing more than the control group students even after the experimental intervention was stopped. Students in the STRAT condition also got significantly higher scores at retention test with an effect size of 0.47 standard deviation, also indicating a continuing higher growth compared to the control group.

No significant effect was found for the STRAT+SA condition. On the contrary, pairwise comparisons of the estimates of the experimental conditions reveal that the STRAT+CA condition also outperforms the STRAT+SA condition with an effect size of 0.29 standard deviation. The difference is however only marginal significant. Furthermore, no interaction effects occurred on the retention test between condition and initial performance level. This implies that the significant positive effect of the experimental STRAT+CA and STRAT conditions is equally strong for initially low scoring children as compared to high achievers.

Table 2.8

*Model Estimates for the Two-Level Analyses of the Fifth-Grade Retention Test Reading Comprehension Achievement Scores*

Parameter	Model			
	0	1	2	3
<i>Fixed</i>				
Intercept	-0.08 (0.11)	-0.07 (0.08)	-0.08 (0.08)	-0.40 (0.12)**
Pretest comprehension		0.69 (0.04)***	0.61 (0.04)***	0.62 (0.04)***
Pretest individual attitude			0.13 (0.04)***	0.13 (0.04)***
Pretest perceived scholastic competence			0.08 (0.04)*	0.08 (0.04)*
STRAT+SA				0.25 (0.19)
STRAT+CA				0.60 (0.19)**
STRAT				0.47 (0.16)**
<i>Random</i>				
Level 2				
$\sigma^2_{\mu 0}$	0.21 (0.08)**	0.12 (0.04)**	0.11 (0.04)**	0.06 (0.03)*
Level 1				
$\sigma^2_{\epsilon 0}$	0.81 (0.06)***	0.42 (0.03)***	0.39 (0.03)***	0.40 (0.03)***
<i>Deviance</i>	1096.16	815.58	778.75	768.97

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

### *Effects on Preoccupation with Attributions and Self-Efficacy Thoughts*

As regards second graders' preoccupation with thoughts relating to self-efficacy in reading or success and failure attributions towards reading, no statistically significant effects were found for any of the experimental conditions.

With regard to fifth graders, an effect of the experimental interventions was found on thoughts relating to failure attributions and negative self-efficacy perceptions towards reading. Table 2.9 and 2.10 summarize the construction of the most appropriate model for respectively the post- and retention test measures. The analyses of the two-level null models (Model 0) show that only the variance at level 1 is significantly different from zero at both the post- and retention test, entailing that the total variance of the dependent variables can be explained by individual differences between students, but not by differences between classes.

As Model 1 in Table 2.9 and 2.10 shows, children's pretest preoccupation with failure attributions and negative self-efficacy perceptions towards reading is a significant predictor of both the post- and retention test measures. Effect sizes are respectively



0.64 and 0.58 standard deviation. Including this pretest measure in the model explains respectively 43% and 33% of the variance between students at post- and retention test. These results indicate that having negative thoughts about one's own reading abilities and associated failure attributions is quite persistent over time in fifth graders.

As it is possible that the variance between students differs according to the pretest measure, we allowed the parameter estimate of this predictor to vary randomly across all students. The results show complex variance at level 1 for both the post- and retention test. More specifically, the variance between students within classes can be considered as a linear function of children's pretest report of thoughts relating to failure attributions and negative self-efficacy perceptions and increases as pretest reports go up. This result implies that the differences in post- and retention test reports between fifth graders within classes are bigger for students who reported more negative thoughts during reading at the beginning of the school year.

As regards other relevant explanatory variables Model 2 in Table 2.9 and 2.10 shows a rather small but significant effect of students' pretest reading comprehension achievement on both the post- and retention test measures. Higher achieving students tend to show a somewhat bigger decrease in having on-line negative thoughts about their own reading abilities. No complex variances at level 1 were found for this significant predictor.

Model 3 in Table 2.9 reveals that by the end of the school year children in the STRAT+CA condition show a significantly bigger relative decrease in being occupied with failure attributions and negative self-efficacy perceptions during reading. This decrease is not only bigger when comparing with the control group (effect size of 0.31), but also when comparing with the STRAT+SA condition (effect size of 0.29).

Table 2.9

*Model Estimates for the Two-Level Analyses of the Fifth-Grade Posttest Measures of Preoccupation with Failure Attributions and Negative Self-Efficacy Perceptions while Reading*

Parameter	Model			
	0	1	2	3
<i>Fixed</i>				
Intercept	0.00 (0.05)	-0.01 (0.04)	-0.01 (0.04)	0.10 (0.08)
Pretest failure attributions and neg. self-efficacy perceptions		0.64 (0.04)***	0.59 (0.04)***	0.58 (0.04)***
Pretest comprehension			-0.13 (0.04)**	-0.14 (0.04)***
STRAT+SA				-0.02(0.11)
STRAT+CA				-0.31 (0.13)**
STRAT				-0.14 (0.10)
<i>Random</i>				
Level 2				
$\sigma^2_{\mu 0}$	0.00 (0.00)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)
Level 1				
$\sigma^2_{\epsilon 0}$	1.00 (0.07)***	0.54 (0.04)***	0.55 (0.04)***	0.55 (0.04)***
$\sigma^2_{\epsilon 0 \text{ attr. \& self-eff.}}$		0.05 (0.02)**	0.06 (0.02)**	0.06 (0.02)**
<i>Deviance</i>	1261.85	993.04	942.45	935.78

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

As can be seen in Table 2.10 the significant effect of the STRAT+CA condition compared to the control group continues to exist at the time of the retention test in sixth grade. Between the end of the school year and the retention test, there was no further relative decrease of negative thoughts during reading, but the significant lower level observed at the posttest persisted. Remarkably, at the retention test a significant positive effect was observed for the STRAT+SA condition as well. So, whereas relative to the control group no decrease in negative thoughts was observed for the STRAT+SA condition between pretest and posttest, such a decrease was observed six months after the intervention was stopped. Consequently, the significant difference between the STRAT+SA and STRAT+CA condition disappears at retention test. Neither at posttest nor at retention test interaction effects occurred between condition and initial report of thoughts relating to failure attributions and negative self-efficacy perceptions while reading on the one hand and between condition and initial performance level on the other hand. This entails that the significant effects are equally strong for low scoring children as compared to high achievers and for children initially reporting much thoughts relating to failure attributions and negative self-efficacy perceptions as compared to students reporting little of these thoughts.

Table 2.10

*Model Estimates for the Two-Level Analyses of the Fifth-Grade Retention Test Measures of Preoccupation with Failure Attributions and Negative Self-Efficacy Perceptions while Reading*

Parameter	Model			
	0	1	2	3
<i>Fixed</i>				
Intercept	0.00 (0.06)	0.03 (0.05)	0.04 (0.05)	0.19 (0.08)*
Pretest failure attributions and neg. self-efficacy perceptions		0.58 (0.04)***	0.49 (0.05)***	0.49 (0.05)***
Pretest comprehension			-0.20 (0.04)*	-0.21 (0.05)***
STRAT+SA				-0.24 (0.12)*
STRAT+CA				-0.31 (0.13)**
STRAT				-0.12 (0.10)
<i>Random</i>				
Level 2				
$\sigma^2_{\mu 0}$	0.02 (0.02)	0.01 (0.02)	0.01 (0.01)	0.00 (0.00)
Level 1				
$\sigma^2_{\epsilon 0}$	0.98 (0.07)***	0.66 (0.05)***	0.65 (0.05)***	0.64 (0.05)***
$\sigma^2_{\epsilon 0 \text{ attr. \& self-eff.}}$		0.04 (0.03)	0.06 (0.02)*	0.06 (0.02)*
<i>Deviance</i>	1133.50	958.94	923.01	916.22

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

It seems obvious that the positive effects of the tutoring activities on fifth-grade students' self-efficacy perceptions is mediated by the positive effects on their reading comprehension achievement. To test this assumption, the residual of respectively the post- and retention test achievement score regressed on the pretest achievement score was added to Model 3 in Table 2.9 and 2.10. Posttest results show a small but significant effect of the added residual score (effect size of 0.18) and simultaneously a slight decrease in the STRAT+CA effect (effect size reduced to 0.25), suggesting that

for the fifth graders involved in cross-age peer tutoring the decrease in negative thoughts can indeed, at least partly, be explained by their growth in reading comprehension. Although on the retention test a decrease in negative self-efficacy related thoughts was found as well for students in both peer tutoring conditions, no such mediating effects were found.

To explore the possible inverse mediating effects of the students' on-line self-efficacy thoughts and associated causal attributions on their growth in reading comprehension Model 3 in the analysis of posttest reading comprehension achievement was extended by adding the posttest measure for self-efficacy perceptions as a predictor. The effect for this latter predictor was small but significant (effect size of 0.08). Simultaneously the effect for the STRAT+CA condition dropped to a nonsignificant level (effect size reduced to 0.28), whereas the effects for the other experimental conditions did not change much. This result suggests that the positive effect of the STRAT+CA condition on reading comprehension achievement is partially mediated by the positive effects on the students' occupation with negative thoughts about their own reading abilities. For the other experimental conditions no such effects were observed. Similar effects were found for the retention test.

## Discussion

The major aim of the intervention study presented in this article was evaluating the effectiveness and endurance of explicit reading strategies instruction, and the surplus value of peer tutoring activities to practice these strategies, as tools for improving second and fifth graders' reading comprehension achievement and self-efficacy judgments toward their reading ability. A pretest posttest retention test design including three experimental conditions and a commensurable control condition was used. In all of the three experimental conditions students got explicit instruction in six reading strategies. In the control condition a traditional approach to reading comprehension was followed, characterized by discussion of text content based on content specific questions to check students' understanding. In one experimental condition reading strategies were practiced in teacher-led whole-class activities. In the other two experimental conditions strategies were practiced in either reciprocal same-age peer tutoring sessions or in cross-age peer tutoring sessions. The study was situated in the challenging context of intact classes to provide a natural setting for the implementation of the interventions.

### *Effects on Reading Comprehension*

With regard to second graders' reading comprehension achievement results show that explicit instruction in reading strategies creates a significant extra learning gain of approximately one quarter of a standard deviation. However it makes no difference whether reading strategies are practiced under the supervision of the class teacher or under the supervision of a fifth grader, who previously received a sound preparation and instruction in relevant tutoring skills and reading strategies. Both conditions create

the same learning gain compared to the control condition. Interestingly, in those two conditions poor readers made as much progress as high achievers. In the long term, six months after the end of the experimental intervention, the positive effect of both the STRAT and STRAT+CA condition disappeared in third grade. Apparently, at this age long lasting effects can only be obtained by continuing the intervention. Future research should try to elucidate this assumption.

Students who practiced the explicitly taught reading strategies in reciprocal same-age peer tutoring dyads did not make extra learning gains compared to students in the control group. This finding parallels a result from Rosenshine and Meister's (1994) review of reciprocal teaching research, revealing significant effects from grades four to adult education and nonsignificant effects for younger - third-grade - students. However, it contrasts with other research applying different age groups, including second graders (Fuchs, Fuchs, Mathes, et al., 1997; Simmons et al., 1995). But with regard to these last mentioned studies it has to be noted that no separate analyses were done on second graders' data. The nonsignificant effect of the STRAT+SA condition leads us to the cautious supposition that the surplus value of explicit instruction in reading strategies, as found for the STRAT condition, is counteracted by practicing the strategies in second-grade same-age dyads. Subsequent research is necessary to verify whether the preparatory tutor training was not appropriate or unsatisfactory for this age group, whether the selected reading strategies were too difficult to practice independently in second-grade reciprocal same-age dyads, or whether same-age peer tutoring is not an appropriate instructional technique for this age group.

For the fifth graders all three experimental conditions appear to create nearly equally big and significant extra learning gains by the end of the school year, with effect sizes between 0.32 and 0.39 standard deviation. Moreover, the three experimental interventions appeared to be as effective for poor readers as for high achievers. These results are congruous with previous research referred to above confirming the positive effect of explicit strategies instruction on children's reading comprehension achievement (Baumann et al., 1992; Block, 1993; Dole et al., 1991; Dole et al., 1996; Pressley et al., 1989). Besides, the differential gains for the experimental students are commensurable with that of reciprocal teaching, one of the most prominent strategy-instruction programs, developed by Palincsar and Brown (1984). Rosenshine and Meister (1994) for example, reported in their review of reciprocal teaching treatments a median effect size of 0.32 when standardized tests were used. The similar effect sizes of the three experimental conditions in the present research leads us to suspect that the way the explicitly taught strategies are practiced - either in teacher-led activities, in reciprocal same-age dyads, or in cross-age dyads - appear to be a matter of secondary importance. In the long term however, six months after breaking off the intervention, only the significant effects of the STRAT and STRAT+CA condition endure. Interestingly, for the STRAT+CA condition the growth continues at approximately the same rate as during the intervention, leading to an effect size of 0.60 standard deviation at retention test. In contrast to this striking result for the cross-age condition, the condition in which students practiced reading strategies in reciprocal same-age dyads does not seem to lead to any significant additional learning growth in the long run. Compared to the results for the posttest, the analyses on the retention test even

show some relapse for the STRAT+SA condition. The long-term effect of the STRAT condition lies in between those for the two peer tutoring conditions. Taken into account these retention test results, the way the strategies are practiced actually do play a considerable part in reading comprehension achievement of upper primary school children.

Comparing the effect sizes of the experimental conditions, it can be concluded that the STRAT+CA condition produces the most advantageous results in the long term. On the average, their learning gains outperform other fifth graders' growth and the growth effectuated in their own tutees. This is all the more striking since in fact during cross-age tutoring fifth-grade tutors are working with texts on their tutees' level, that is second-grade level instead of fifth-grade level, and their "only" task is to supervise their tutees' reading process, to monitor their tutees' understanding and to try to enhance comprehension by having their tutees apply the reading strategies. Still they effectuated the largest long term progress in reading comprehension. One possible explanation relates to the nature of the tutoring task in connection to the possibility to practice metacognitive skills. The tutoring task requires tutors to focus their attention on the tutee's reading. The fact not having to read themselves may make it easier for tutors to give full attention to the metacognitive skills involved in monitoring and regulating the reading process. In other words: it may be easier and more appropriate to practice the metacognitive skills involved by monitoring and regulating someone else's reading process compared to one's own reading process, more particularly if this someone else is a younger student with a lower reading level. In that respect cross-age tutoring can be seen as a very powerful learning environment for the improvement of the tutors' metacognitive skills, leading to a long term improvement of their own reading comprehension achievement, even without having done much reading themselves. Of course the interpretation of the results presented here, needs to be confirmed in subsequent research in which a more detailed analysis of the interactions during the peer tutoring sessions is performed.

A possible explanation for the disappointing long-term results for the STRAT+SA condition can be found in the findings from some non-systematic observations, indicating a gradual fading away of the reciprocal tutor and tutee roles, less academic responding, a decrease in time-on-task, and less highly occupied activity in the same-age dyads compared to the cross-age dyads. However, additional more qualitatively oriented analyses of the functioning and interactions in the reading dyads are requisite to give a decisive answer about this assumption. Especially the blurred reciprocal tutor and tutee roles merit particular attention, for prior research reported that same-age partners engaging in role reciprocity made greater reading gains than students in dyads with fixed tutor and tutee roles (Simmons, Fuchs, Fuchs, Hodge, & Mathes, 1994).

### *Effects on Preoccupation with Attributions and Self-Efficacy Thoughts*

Our finding that effects were only found on students' negative thoughts about their own reading proficiency and with regard to their failure attributions and that no effects at all were found on students' positive thoughts and success attributions is in line with findings from Hiebert, Winograd, and Danner (1984), Marsh (1984), and Marsh et al.

(1984) that high achievers and low achievers appear to differ more from each other in their failure attributions, with low achievers attributing failures more to lack of ability, than in their success attributions.

For second graders we did not find any effects on being preoccupied with self-efficacy related thoughts whatsoever. This does not necessarily mean that supplying reading strategies instruction or organizing peer tutoring does not have any effect on second graders' self-efficacy perceptions with regard to reading. Maybe in the case of second graders such effects do not reveal themselves in students' preoccupations with self-related thoughts. Or maybe our way of assessing such preoccupations is not apt for second graders. One should take into account that altering children's self-efficacy perceptions can not be easily achieved. Therefore, subsequent research should study whether extending the intervention succeeds in producing significant effects.

As regards fifth graders' self-efficacy related thoughts an interesting result was found, revealing that by the end of the school year children engaged in cross-age tutoring activities are significantly less occupied with failure attributions and negative self-efficacy related thoughts compared to students in the control and students in the STRAT+SA condition. Moreover, the significant difference with the control group endured at retention test. As it can be assumed that high scores for thoughts relating to failure attributions and negative self-efficacy perceptions indicate the presence of low self-esteem with regard to reading ability, these results indicate that opportunities to take on the role of tutor in cross-age dyads have a positive impact on fifth graders' self-esteem and perceived competence.

Unlike the results of the STRAT+CA condition, the effects of the STRAT+SA condition are not univocal. By the end of the school year children report significantly more negative self-efficacy perceptions and thoughts relating to failure attributions than their peers in the STRAT+CA condition. However, six months later at retention test they report significantly less negative thoughts than the control group and the significant difference as compared to the STRAT+CA condition disappears. Clearly, additional research is necessary to shed light on the ambiguous results of the STRAT+SA condition and to ensure that the positive results of the STRAT+CA condition in the present study are replicable.

### *Conclusion*

In sum, in accordance with previous research the present study generally corroborates the first hypothesis and documents the feasibility of fostering reading comprehension achievement in second and fifth graders of different ability levels by supplying them with explicit instruction in reading strategies. Contrary to most other research (e.g., Alfassi, 1998; De Corte et al., 2001; Walraven, 1995), the significant results are found on standardized reading tests. These findings support the current cognitively based views of strategic reading comprehension instruction and demonstrate the educational benefits of explicit reading strategies instruction in intact classrooms as part of the

overall curriculum. However, only for fifth graders long-term effects can be documented.

The second hypothesis regarding the surplus value of practicing the application of reading strategies in tutoring dyads as compared to practice in teacher-led activities, can not really be confirmed. Neither for second nor for fifth graders significant differences were found between the STRAT condition on the one hand and the STRAT+SA or STRAT+CA condition on the other hand. Although the difference is not significant, attention should be drawn, however, to the effect size of the STRAT+CA condition at the retention test, indicating a somewhat stronger long-term effect as compared to the STRAT condition.

As to the third and fourth hypothesis, it was assumed that STRAT+CA would be more effective than STRAT+SA in fostering second and fifth graders' reading comprehension achievement. Since in second grade the STRAT+CA condition significantly outperformed the control condition at posttest, whereas the STRAT+SA condition was not significantly different from this reference group, the results point in this direction. Comparing both tutoring conditions with each other does however not reveal mutually significant differences. Likewise in fifth grade, at the end of the school year no significant differences were found between the same-age tutoring condition and the cross-age tutoring condition. But six months later the difference became marginally significant, which suggests the correctness of the fourth hypothesis.

With regard to the effects of peer tutoring activities on students' preoccupation during reading with self-efficacy related thoughts, a general finding was that only effects with regard to negative thoughts were found. The decrease in such thoughts that was observed in some conditions indicates that students became more confident in their own reading competence and were less preoccupied with doubts. This was particularly true six months after finishing the intervention for fifth graders who had been engaged in cross-age peer tutoring activities. So, for this condition and measurement occasion the fifth hypothesis can be confirmed.

Since at the end of the school year children engaged in cross-age tutoring activities were also significantly less occupied with negative self-efficacy related thoughts as compared to students in the reciprocal same-age peer tutoring condition, the seventh hypothesis is corroborated for the posttest measurement occasion as well. This does however no longer apply to the retention test results. For second grade neither the fifth nor the sixth hypothesis could be confirmed.

By way of conclusion, some limitations of the study should be mentioned. A first comment can be made on some of the measurement instruments used. The fact that the present study reported significant effects on second and fifth graders' standardized reading comprehension tests instead of on experimenter-designed tests, ensures the strength of the experimental intervention. As pointed out in the discussion section, it is quite reasonable that more particularly the long-term effects of the cross-age peer tutoring condition on fifth graders' reading comprehension achievement can be attributed to an improvement in their metacognitive skills and use of reading comprehension strategies. It would explain how fifth graders can make this considerable improvement in reading comprehension at their own level while in fact they only have been monitoring and regulating second graders' reading of second-

grade level texts. However, no direct measurement of the use of reading strategies nor metacognitive activity has been realized in this study. Although an attempt was made to organize such a measurement by means of questionnaires, we have not been successful in establishing valid measures. So, in future research assessing students' application of reading strategies should merit particular attention. Using thinking-aloud protocols could shed light on students' reading behavior, application of reading strategies, and metacognitive regulation of the reading process. These supplementary sources of information would enable us to ascertain whether the impact of the interventions on standardized reading comprehension tests can indeed be attributed to an increasing mastery of reading strategies. Moreover, using thinking-aloud protocols could also document the validity of the questionnaire with regard to preoccupation with negative self-efficacy related thoughts and enable us to assess students' tendency to social desirability in answering the questionnaire.

As mentioned above, a second limitation of the study relates to the absence of detailed in-depth observations of the quality of students' interactions in cross-age and reciprocal same-age dyads.

Thirdly, the design of the present study enables us to analyze the differential effects of cross-age and reciprocal same-age peer tutoring activities to practice the application of reading strategies. But due to the complexity of the intervention, it is impossible to determine the relative contribution of each of the components of the intervention (e.g., the separate reading strategies, the gradual transfer of responsibility in applying the strategies from teacher to students, ...). A profound component analysis could obviate this drawback and clarify essential and dispensable components of the intervention. However, as Brown et al. (1996) argue, the approach of the present research is appropriate and defensible when the interest is in evaluating the efficacy of a multicomponential instructional package.

Finally, the extensive assistance and coaching of the teachers provided by the first author, may also be seen as a burden on the opportunities for implementing the innovation in instructional practice and restricts our capacity to generalize results to other situations where help of the sort we provided is absent. Given that it is unlikely that regular in-service training could have comparable levels of support available for teachers, the transfer of the present findings is uncertain. Therefore, we are currently monitoring the effects of a more restricted training of teachers in an attempt to replicate the positive effects of the present study.

## References

- Aarnoutse, C. A. J. (1995). *Probleemidentificatie en aanzet voor een actieplan taal* [Problem identification and initiative towards an action plan for language]. Nijmegen, The Netherlands: Vakgroep Onderwijskunde.
- Aarnoutse, C. A. J. (1996). *Leesattitudeschaal* [Reading attitude scale]. Nijmegen, The Netherlands: Berkhout.
- Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal*, 35, 309-332.



- Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.
- Alvermann, D. E. (2000). Classroom talk about texts: Is it dear, cheap or a bargain at any price? In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 170-192). New York: Teachers College Press.
- Ames, C. (1984). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Educational Psychology*, 76, 478-487.
- Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353-394). New York: Longman.
- Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior: A Journal of Literacy*, 24, 143-172.
- Bentz, J. L., & Fuchs, L. S. (1996). Improving peers' helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly*, 19, 202-215.
- Block, C. C. (1993). Strategy instruction in a literature-based reading program. *The Elementary School Journal*, 94, 139-151.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.
- Brus, B. Th. (1969). *Eén-minuut-test* [One minute test]. Nijmegen, The Netherlands: Berkhout.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Casteel, C. P., Isom, B. A., & Jordan, K. F. (2000). Creating confident and competent readers. Transactional strategies instruction. *Intervention in School and Clinic*, 36, 67-75.
- Chard, D. J., Simmons, D. C., & Kameenui, E. J. (1998). Word recognition: Research bases. In D. C. Simmons, & E. J. Kameenui, (Eds.), *What reading research tells us about children with diverse learning needs. Bases and basics* (pp. 141-167). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.
- De Corte, E., Verschaffel, L., & Van De Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Educational Psychology*, 71, 531-559.
- Dole, J. A. (2000). Explicit and implicit instruction in comprehension. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 52-69). New York: Teachers College Press.
- Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly*, 31, 62-88.

- Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61, 239-264.
- Durkin, D. (1978-1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14, 481-533.
- Ezell, H. K., & Kohler, F. W. (1992). Use of peer-assisted procedures to teach QAR reading comprehension strategies to third-grade children. *Education and Treatment of Children*, 15, 205-228.
- Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on self-concept and mathematics achievement. *Journal of Educational Psychology*, 87, 272-281.
- Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology*, 84, 331-339.
- Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *The Journal of School Psychology*, 28, 309-323.
- Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81, 173-177.
- Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, C. L. (1994). The nature of students interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75-103.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K. & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 97, 223-249.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [Step out! Manual]. Utrecht, The Netherlands: Sardes.
- Greenwood, C. R., Carta, J. C., & Hall, R. V. (1988). The use of peer tutoring strategies in classroom management and educational instruction. *School Psychology Review*, 17, 258-275.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.
- Harter, S. (1985). *Manual for the self-perception profile for children*. Denver, CO: University of Denver.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 33-68). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Henk, W. A., & Melnick, S. A. (1992). The initial development of a scale to measure "perception of self as reader". In C. K. Kinzer, & D. J. Leu (Eds.), *Literacy research, theory, and practice: Views from many perspectives* (pp. 111-117). Chicago: National Reading Conference.

- Henk, W. A., & Melnick, S. A. (1995). The Reader Self-Perception Scale (RSPS): A new tool for measuring how children feel about themselves as readers. *Reading Teacher*, 48, 470-482.
- Hiebert, E. H., Winograd, P. N., & Danner, F. W. (1984). Children's attributions for failure and success in different aspects of reading. *Journal of Educational Psychology*, 76, 909-919.
- Hox, J. J. (1994). *Applied multilevel analysis*. Amsterdam: TT-publikaties.
- Hox, J. J., & Kreft, I. (1994). Multilevel analysis methods. *Sociological Methods and Research*, 22, 283-299.
- Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: verbal versus visual strategies. *Journal of Educational Psychology*, 92, 772-782.
- Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, 96, 275-293.
- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, 99, 3-21.
- Lambiotte, J. G., Dansereau, D. F., O'Donnell, A. M., Young, M., D., Skaggs, L. P., Hall, R. H., & Rocklin, T. R. (1987). Manipulating cooperative scripts for teaching and learning. *Journal of Educational Psychology*, 79, 424-430.
- Lories, G., Dardenne, B., & Yzerbyt, V. Y. (1998). From social cognition to metacognition. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition. Cognitive and social dimensions* (pp. 1-15). London: Sage.
- Marsh, H. W. (1984). Relations among dimensions of self-attribution, dimensions of self-concepts and academic achievements. *Journal of Educational Psychology*, 76, 1291-1308.
- Marsh, H. W., Cairns, L., Relich, J., Barnes, J., & Debus, R. L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. *Journal of Educational Psychology*, 76, 3-32.
- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.
- Mathes, P. G., Torgesen, J. K., Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal*, 38, 371-410.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Palincsar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. *Remedial and Special Education*, 12, 43-53.
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 1239-1252.
- Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition and motivation. *Developmental Review*, 6, 25-56.

- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Vol. 2* (p. 609-640). New York: Longman.
- Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Vol. 2* (pp. 815-860). New York: Longman.
- Pressley, M. (2000). Comprehension instruction in elementary school: A quarter-century of research progress. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 32-51). New York: Teachers College Press.
- Pressley, M., & Allington, R. (1999). What should reading instructional research be the research of? *Issues in Education*, 5, 1-35.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92, 513-555.
- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal*, 90, 3-32.
- Pressley, M., Wharton-McDonald, R., Hampston, J. M., & Echevarria, M. (1998). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies of Reading*, 2, 159-191.
- Raphael, T. E. (2000). Balancing literature and instruction: Lessons from the Book Club Project. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for Meaning. Fostering Comprehension in the Middle Grades* (pp. 70-94). New York: Teachers College Press.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, 64, 479-530.
- Roswal, G. M., Mims, A. A., Evans, M. D., Smith, B., Young, M., Burch, M., Croce, R., Horvat, M. A., & Block, M. (1995). Effects of collaborative peer tutoring on urban seventh graders. *Journal of Educational Research*, 88, 275-279.
- Simmons, D. C., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice*, 9, 203-212.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Staphorsius, G., & Krom, R. (1996). *Toetsen Begrijpend Lezen* [Reading Comprehension Tests]. Arnhem, The Netherlands: Cito.
- Topping, K. J. (1988). *The peer tutoring handbook: Promoting cooperative learning*. London: Croom Helm.
- Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.

- Topping, K. J. (2001). *Thinking Reading Writing. A practical guide to paired learning with peers, parents and volunteers*. London: Continuum.
- Utley, C. A., Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children*, 29, 69-92.
- Van Den Broek, P., & Kremer, K. E. (2000). The mind in action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 1-31). New York: Teachers College Press.
- Veerman, J. W., Straathof, M. A. E., Treffers, Ph. D. A., Van den Bergh, B. R. H., & ten Brink, L. T. (1997). *Competentiebelevingsschaal voor Kinderen (CBSK)* [Self-Perception Profile for Children]. Lisse, The Netherlands: Swets & Zeitlinger.
- Verhoeven, L. (1993). *Lezen met begrip 1* [Reading with comprehension 1]. Arnhem, The Netherlands: Cito.
- Walraven, M. (1995). *Instructie in leesstrategieën. Problemen met begrijpend lezen en het effect van instructie aan zwakke lezers* [Instruction in reading strategies. Difficulties with reading comprehension and the effect of instruction for poor readers]. Amsterdam: Paedologisch Instituut.
- Weisberg, R. (1988). 1980s: A change in focus of reading comprehension research: A review of reading/learning disabilities research based on an interactive model of reading. *Learning Disability Quarterly*, 11, 149-159.
- Weterings, A. C. E. M., & Aarnoutse, C. A. J. (1986). De praktijk van het onderwijs in begrijpend lezen [Instruction in reading comprehension in practice]. *Pedagogische Studiën*, 63, 387-400.

## Chapter 3

### Reanalyzing the first study's effects on fifth graders' reading comprehension: A three-level repeated measures approach\*

#### Abstract

**Background.** Explicit reading strategies instruction and engaging students in interaction about texts show great promise in promoting students' reading comprehension ability. The present intervention study combines both aspects.

**Aims.** The study examines the educational benefits of explicit reading strategies instruction, followed by practice in (a) teacher-led whole-class activities (STRAT), (b) reciprocal same-age (STRAT+SA), or (c) cross-age peer tutoring activities (STRAT+CA) on fifth graders' reading comprehension achievement.

**Sample.** Twenty-two fifth-grade teachers and their 454 students from 19 different schools throughout Flanders (Belgium) participated.

**Method.** A quasi-experimental pretest posttest retention test design was used with three experimental (STRAT, STRAT+SA, and STRAT+CA) and a matched control group. The experimental interventions were implemented during an entire school year.

**Results.** Multilevel analysis revealed that the STRAT and STRAT+CA condition made a significantly larger progress from pretest to retention test than the control group. The significant major progress was especially situated from pretest to posttest, during which the intervention took place. Concerning the STRAT+SA condition no significant differences with regard to the control group were detected. Pairwise comparisons of the experimental conditions indicated that the STRAT+CA condition's progress from pretest to retention test also exceeded the STRAT+SA condition's advancement significantly.

**Conclusion.** The study corroborated the efficacy of the STRAT and STRAT+CA condition's interventions as feasible tools to enhance fifth graders' reading comprehension achievement.

Moreover, a continued growth till at least six months after finishing the program was shown.

#### Introduction

Reading holds an important place in our modern society. In daily life of both children and adults written language is appealed to at all kinds of places and times. Accordingly, it is indispensable that one has satisfactory functional literacy at his disposal (Walraven, 1995). This entails that learning to read is one of the most crucial learning processes children are involved in at primary school. Nevertheless, learning to read is not that easy for everyone. Successful reading requires many basic skills on which children can drop out, such as the identification of letters and letter combinations, the mapping of letters onto sounds, the synthesizing of sounds, and the recognition of words and syntax. However, as Van Den Broek and Kremer (2000) argue,

---

\* Based on: Van Keer, H. Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. The present chapter is submitted for publication in the *British Journal of Educational Psychology*.

the ultimate goal is for readers to *learn* from text: to recognize the depicted facts or events, to connect them to each other and to prior knowledge, and to memorize the results so that they can be used later. This goal requires additional higher-order processes, such as inference making and reasoning. (p. 1)

It is especially this ultimate goal of reading comprehension that many children seem to have huge problems with.

This introductory section further presents a definition of reading comprehension and the commonly recognized sources of comprehension failure, namely reader, text, and context properties. The lack of explicit instruction in reading strategies as an additional possible cause of poor reading comprehension is expounded next. Subsequently, research evidence is discussed demonstrating the importance of collaboration and interaction among peers to enhance reading comprehension. Attention is especially called to the prospects of peer tutoring. Finally, we go into the aim of the present study.

### *Reading Comprehension and Sources of Comprehension Failure*

Reading comprehension can be defined as constructing a mental representation of textual information and its interpretation (Van Den Broek & Kremer, 2000) or, in other words, as extracting meaning from written information (Aarnoutse, 1991). Notwithstanding the simplicity of this definition it is a complex process, in which different components affect each other reciprocally. Because reading comprehension can be seen as a continuous interaction between the reader and the text (Paris, Wasik, & Turner, 1991), it is commonly accepted that both reader characteristics and text properties are main sources of failure in comprehension.

Significant reader characteristics are the general cognitive abilities, the decoding skills, the amount of relevant background knowledge, the vocabulary, the motivation, and the interest of the reader. Relevant text properties are the difficulty, the length, and the structure of the text, as well as the predictability of the information, and the clarity of the content (Palincsar & Brown, 1984; Van Den Broek & Kremer, 2000; Walraven, 1995). In addition to reader and text characteristics, the specific context in which reading occurs also affects the probability of successful comprehension. Consequently, it may be argued that reading comprehension suffers if any aspect of the interaction among the reader, text, and immediate context goes awry (Van Den Broek & Kremer, 2000). Nevertheless, instruction in the requisite basic and higher-order reading skills, or lack thereof, can influence children's performances as well, for not all students can be expected to become independently competent with the skills used by proficient readers (Hartman, 2001).

*Need of Explicit Instruction in Reading Comprehension Strategies*

It is known that reading comprehension requires more than being able to decode correctly. Cognitively based views of reading comprehension emphasize that good "comprehenders" do much more than just word-, phrase-, or sentence-level processing; they make use of a flexible repertoire of comprehension-fostering and comprehension-monitoring activities (Dole, Duffy, Roehler, & Pearson, 1991; Palincsar & Brown, 1984). They have a purpose for reading, they monitor comprehension while reading, and they reflect on their reading process. Moreover, they have a thorough command of a whole range of appropriate strategies to handle comprehension failures. These strategies are conscious, instantiated, and flexible plans readers apply and adapt deliberately to a variety of texts and tasks (Baker & Brown, 1984; Paris et al., 1991; Pressley & Allington, 1999; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989). Summarizing, it may be stated that proficient readers are typified by the mastery and use of both metacognitive and cognitive strategies that facilitate text comprehension.

Cognitive strategies can be defined as mental and behavioral activities, such as rereading, activating prior knowledge, and adjusting reading speed, used to increase the likelihood of comprehending (Van Den Broek & Kremer, 2000). Metacognitive strategies are self-monitoring and self-regulating activities, focusing on the process and product of reading. They enhance readers' awareness of whether or not they comprehend what they are reading, their ability to judge the cognitive demands of a task, and their knowledge of when and how to employ a specific cognitive strategy as a function of text difficulty, situational constraints, and the reader's own cognitive abilities (Baker & Brown, 1984; Dole, 2000; Duffy et al., 1987; Ehrlich, Kurtzcostes, & Lorient, 1993; Gourgey, 2001; Lories, Dardenne, & Yzerbyt, 1998; Van Den Broek & Kremer, 2000; Van Kraayenoord & Schneider, 1999; Weisberg, 1988).

Notwithstanding the importance of cognitive and metacognitive strategies as tools for enhancing reading comprehension, there is no reason to assume that all elementary students spontaneously discover them and appeal to strategic processes when confronting texts that are challenging to comprehend (Hartman, 2001; Pressley & Allington, 1999). Nevertheless, research results reveal that even when children do not use effective comprehension strategies on their own, explicit reading strategies instruction is a feasible tool for teaching students to apply them successfully (Brand-Gruwel, Aarnoutse, & Van Den Bos, 1998; De Corte, Verschaffel, & Van de Ven, 2001; Dole et al., 1991; Duffy et al., 1987; Haller, Child, & Walberg, 1988; Pressley, 2000; Pressley et al., 1989; Walraven, 1995). Explicit instruction refers in this case to the purposive activities of a teacher to make children fully aware of the active character of the reading process and of the importance of comprehension-fostering and monitoring activities (Paris et al., 1991). This implies that children are instructed in strategies which support comprehension; in why, where, and when to use them; as well as in how to adapt them to various situations (Hartman, 2001; Paris, Lipson, & Wixson, 1983; Paris et al., 1991; Pressley, 2000).

Unfortunately, research shows that the instructional practice of teaching reading in primary schools is still very traditional, characterized by a great deal of comprehension "testing" (questioning students about the content of a text after reading it), and hardly



any explicit and continuous comprehension instruction that aims at making students astute in strategy selection, use, and evaluation (Aarnoutse, 1995; Aarnoutse & Weterings, 1995; Dole, 2000; Paris & Oka, 1986; Pressley, Wharton-McDonald, Hampston, & Echevarria, 1998; Weterings & Aarnoutse, 1986).

### *Need of Peer-Led Interaction about Texts*

Besides the importance of explicit reading strategies instruction, there is evidence that opportunities to participate in peer-led interaction on structured reading activities also make up an important part of reading instruction that aims at an actual increase in comprehension, higher-level cognition and the application of self-regulation strategies (Alfassi, 1998; Almasi, 1996; Almasi, McKeown, & Beck, 1996; Fuchs & Fuchs, 2000; Fuchs, Fuchs, Mathes, & Simmons, 1997; Greenwood, Delquadri, & Hall, 1989; Guthrie, Schafer, Wang, & Afflerbach, 1993; Johnson-Glenberg, 2000; Klingner, Vaughn, & Schumm, 1998; Mathes & Fuchs, 1994; Mathes, Torgesen, & Allor, 2001; Morrow & Smith, 1990; Rosenshine & Meister, 1994; Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995). With respect to reading comprehension, the traditional classroom interaction pattern of teacher question - student response - teacher evaluation seems insufficient for the development of the deeper meaning of texts (Cazden, 1986). Relying on the teacher to serve as the interpretive authority may cause students to become passive learners. Conversely, in order for children to become self-regulated readers and thinkers, they need to take an active role and to recognize and resolve their own discrepancies with texts (Almasi, 1996; Gourgey, 2001). Research demonstrates that this active reading behavior is promoted by providing students opportunities to engage in peer-led interaction about texts. More particularly it is shown that interaction between peers encourages children to talk not only about what they are reading, but also about what they do when reading (McCarthy, Hoffman, & Galda, 1999). Through discussions, peer conferences, peer tutoring, and cooperative activities students implement, evaluate and modify strategy acquisition and use, and discuss the application of a strategy in situations other than the reading lessons (Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984). Moreover, discussions between peers provide opportunities for metacognitive exchanges and modeling (Palincsar, David, Winn, & Stevens, 1991). In this way, children's knowledge about reading and reading strategies, as well as their ability to apply relevant strategies increases when hearing others talk about their reading process.

Given the important linkages between peer interaction about text on the one hand and improved reading comprehension, developing strategies for comprehending text, as well as applying self-regulation strategies on the other hand, one might expect to find teachers using a fair amount of peer interaction and discussion with regard to reading comprehension instruction. In reality however, this does not seem to be the case. According to Alvermann (2000) student-centered discussion is anything but common practice in most classrooms. Nonetheless, researchers have explored a number of specific ways to encourage social and task-oriented interaction among peers in the classroom. A well-known study in this respect is the research of Palincsar and Brown

(1984) on reciprocal teaching of comprehension strategies in the context of small groups. The teacher initially explains and models the strategies, but gradually transfers responsibility to the members of the group, with individual students taking turns leading the reading group in use of the strategies. Other representatives in the area of strategy learning and teaching are Pressley and his colleagues (Brown, Pressley, Van Meter, & Schuder, 1996; Pressley et al., 1992) with their work on transactional instruction of comprehension strategies in reading groups. Contrary to the previous research with attention to interaction in small groups, the present study takes an interest in "peer tutoring" in dyads, for studies of paired peer tutoring in reading and thinking skills are relatively rare (Topping, 2001). With the exception of research on Peer-Assisted Learning Strategies (PALS) (Fuchs, Fuchs, Mathes, et al., 1997) almost all of the research on tutoring students in reading has involved word-level oral reading or low-level comprehension activities rather than comprehension strategy training (Pearson & Fielding, 1991).

### *Peer Tutoring*

Peer tutoring can be defined as "people from similar social groupings who are not professional teachers helping each other to learn, and learning themselves by teaching" (Topping, 1996, p. 322). This means that the term peer tutoring covers a whole series of practices that employ peers as one-on-one teachers to provide individualized instruction, practice, repetition, and clarification of concepts (Topping, 1988; Utley & Mortweet, 1997). Peer tutoring is characterized by specific role taking: at any point someone has the job of tutor, while the other is in a role as tutee (Topping, 1996). Two large categories of peer tutoring can be distinguished. Children can be paired with other children from within their own classroom. This variant is called same-age tutoring. The specific form of same-age tutoring in which the students alternate on a regular basis between the tutor and tutee role is called reciprocal same-age tutoring (Fantuzzo, King, & Heller, 1992). The second variant is called cross-age tutoring and refers to older students tutoring younger students. In this case role reciprocity is basically not applicable.

As compared to conventional teacher-mediated instruction, peer tutoring increases individualization, time on task, immediate and specific feedback, reinforcement and error correction, as well as opportunities to respond, academic engagement and relevant academic behaviors that are related to specific academic tasks (Greenwood & Delquadri, 1995; King-Sears & Bradley, 1995; Utley & Mortweet, 1997). Significant to mention is that the teacher remains important. However, he takes on a new role of monitor, coach and facilitator. The teacher must be active, needs to check in with groups, must monitor progress and interaction and must provide feedback. When necessary, the teacher can model good helping behavior, give examples, and observe whether targeted skills are being practiced (Webb & Farivar, 1999). Different from other formal and informal occasions of helping behavior among students, peer tutoring is embedded structurally in the curriculum and classroom organization. Moreover, the tutoring activities are preceded by a special preparatory instruction, for research shows that peer tutoring is less effective when no attention is paid to a sound prior training of

the tutors, related to the content involved and to necessary social and communication skills (e.g., Bentz & Fuchs, 1996; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997).

Over the last two decades, both same-age and cross-age peer tutoring have been proven successful in various circumstances. It has been implemented in a variety of curriculum areas and at different age levels. Research results indicate positive effects on academic achievement, for both tutor and tutee (e.g., Cohen, Kulik, & Kulik, 1982; Fantuzzo, Polite, & Grayson, 1990; Fantuzzo et al., 1992; Greenwood, Terry, Arreaga-Mayer, & Finney, 1992; Greenwood, Terry, Utley, Montagna, & Walker, 1993), in particular for children with learning problems (e.g., Bentz & Fuchs, 1996; Fuchs, Fuchs, Mathes, et al., 1997; Klingner & Vaughn, 1996; Mathes & Fuchs, 1994). Positive effects have also been found on tutors' and tutees' social and emotional functioning, more specifically with regard to improved self-efficacy perceptions, more positive self-concepts and social relationships, and better attitudes towards the curriculum areas treated in the tutoring sessions (e.g., Cohen et al., 1982; Fantuzzo et al., 1992; Fantuzzo, Davis, & Ginsburg, 1995; Greenwood, Carta, & Hall, 1988; Mathes & Fuchs, 1994; Roswal et al., 1995).

A striking research result is that tutors seem to benefit even more from tutoring than the students receiving the individual tuition (Bargh & Schul, 1980; Fitz-Gibbon, 1988; Greenwood et al., 1988; Lambiotte et al., 1987). This result confirms the assumption underlying the cross-age approach that older children can benefit from tutoring experiences. With regard to the improved academic achievement this can be explained by the fact that tutors are challenged to consider the subject fully from different perspectives, to engage in active monitoring to identify and correct errors, to reorganize and clarify their own knowledge and understandings, and to elaborate on information in their explanations in order to provide help that can be useful to the tutee (Fuchs & Fuchs, 2000; Wittrock, 1990 in Terwel, Gillies, van den Eeden, & Hoek, 1999). According to another hypothesis, the opportunity to act as tutor and the status associated with the tutoring role may increase self-perceptions of competence enough to improve academic performance (Bierman & Furman, 1981). Taking into account both explanations concerning the metacognitive and cognitive demands associated with the tutoring role on the one side and the status associated with that role on the other side, it seems recommendable to adopt role reciprocity in the case of same-age tutoring. In this way, both students can profit by the benefits connected to the tutor role. This assumption was confirmed by research reporting that partners engaging in role reciprocity made greater reading gains than partners who did not (Simmons, Fuchs, Fuchs, Hodge, & Mathes, 1994).

### *Study's Purpose*

Explicit reading strategies instruction, as well as engaging students in interaction about texts shows great promise in promoting and increasing students' reading comprehension ability. In view of combining both research fields an intervention study was designed, intended to evaluate the efficacy of complex reading comprehension

instructional innovations in real classrooms. More specifically we wanted to provide information on the differential effects of explicit reading strategies instruction, followed by practice in (a) teacher-led whole-class activities, (b) reciprocal same-age peer tutoring activities, or (c) cross-age peer tutoring activities.

By linking reading strategies instruction with peer tutoring, the present study builds on former research, focusing on peer-led interaction and reading strategies acquisition (Brown et al., 1996; Fuchs, Fuchs, Mathes, et al., 1997; Palincsar & Brown, 1984; Pressley et al., 1992). Moreover the study extends prior research by (a) sampling a relatively large number of participants from 22 classrooms in 20 schools throughout Flanders (Belgium); (b) supporting teachers to implement the innovations in their regular classroom context with the participation of all students in the course of an entire school year; (c) including long-term maintenance assessments; (d) using standardized reading comprehension tests not linked directly to the treatment; (e) applying multilevel modeling to take the hierarchical nesting of students in classes into account; (f) and enabling an explicit comparison of the differential impact of cross-age and reciprocal same-age peer tutoring within the same study. This is unique with respect to other peer tutoring studies, merely focused on studying the effectiveness of one peer tutoring variant. So far, there is only indirect evidence that cross-age tutoring is more effective than same-age tutoring, for the meta-analysis of Cohen et al. (1982) reveals larger effect sizes for both tutors' and tutees' achievement in cross-age tutoring programs. However, no statistically significant differences between the variants were found.

In addition to contribute to theory building, the present study intends to bridge the aforementioned gap between instructional practice and research evidence indicating the possibilities of explicit reading strategies instruction and peer interaction about texts. Therefore, the study was conceived as a "design experiment", which meets the twofold goal of advancing theory building and optimizing classroom education by creating and evaluating complex instructional interventions, that embody the present understanding of effective learning processes and powerful learning environments, in real classrooms and in partnership between researchers and practitioners (De Corte, 2000).

## **Method**

### *Design*

To investigate the differential effects of explicit reading strategies instruction, followed by practice in teacher-led whole-class activities, reciprocal same-age, or cross-age peer tutoring activities, a pretest posttest retention test control group design was used. More specifically, classes were assigned to four conditions: (a) explicit reading strategies instruction and practicing the topics during teacher-led whole-class activities (STRAT), (b) explicit reading strategies instruction and practicing the topics in class-wide reciprocal same-age dyads (STRAT+SA), (c) explicit reading strategies instruction and practicing the topics in class-wide cross-age dyads (STRAT+CA), and

(d) traditional reading comprehension instruction without explicit instruction or peer tutoring (control group). The effects of peer tutoring activities independent of teacher-directed explicit reading strategies instruction were not investigated, because in traditional reading comprehension instruction no explicit instruction is provided, so there is no content for tutoring.

Because complete naturally constituted classes were assigned to the conditions in the design experiment, the research took place in an ecologically valid setting and the design can be referred to as quasi-experimental. This setting provides a more stringent test of the successful implementation of the interventions than studies in tightly controlled laboratory settings, of which research results can not simply be transferred to the context of real-life classrooms. To the extent that the results show that the interventions are effective in the quasi-experimental context, they will strengthen their applicability and effectiveness in natural classroom settings.

### *Participants*

Twenty-two fifth-grade teachers and their 454 students from 19 different schools throughout Flanders were involved in the study. Except for one inner-city school with a predominantly low SES and ethnic minority population, all classes had a mainly white, Flemish population. The majority of the children were from middle-class families. Classes are to be considered as academically heterogeneous. The age of the students ranged from 9 to 12 years, with an average of 10 years and 5 months at the beginning of the school year. There was approximately an even gender distribution: 220 students were boys and 234 students were girls. The number of students in a class ranged from 10 to 30, with an average of almost 22 ( $SD = 5.00$ ) students per class, which is quite representative for the Flemish situation. The majority of the students (422) were native Dutch speakers, which is the medium of instruction in Flanders. Teachers were on average 42 years old and had 20 years of teaching experience. Five of the 22 fifth-grade teachers were men. None of them had experience in the experimental interventions.

Participating teachers were selected from a group of 40 fifth-grade teachers, who declared oneself willing to be engaged in a long-term research study. The selection was based on the geographical distribution of the schools throughout the whole of Flanders on the one hand, and on the possibility to match teachers and classes as closely as possible with regard to teachers' teaching experience, class size, students' age, gender distribution, and dominating mother tongue on the other hand.

Teachers were randomly assigned to the experimental STRAT or tutoring conditions. Within the peer tutoring conditions they were allowed to opt in favor of the STRAT+SA or STRAT+CA condition according to their own preference and the readiness of a second-grade colleague to participate in the STRAT+CA condition. Control group classes were selected to match the teachers and class groups in the experimental conditions. More specifically, 8 teachers and 177 students participated in the STRAT condition, 4 teachers and 101 students in the STRAT+SA condition, 4 teachers and 69 students in the STRAT+CA condition, and 6 teachers and 107 students

served as the control group. The fact that twice the number of classes were assigned to the STRAT condition as compared to the STRAT+SA and STRAT+CA condition had to do with the continuation of the present study, intended to replicate the implementation of both tutoring conditions with a larger number of classes, including both the STRAT, STRAT+CA and STRAT+SA classes.

## *Intervention*

### *Components of Teacher Support*

Because the interventions in the three experimental conditions (STRAT, STRAT+SA, and STRAT+CA) were not implemented by the researcher but by the regular classroom teachers, these teachers were prepared and supported by the author of the article to ensure that they were administering the innovations with fidelity. To that end, teachers in all experimental conditions were provided with a specific elaborated manual including all materials necessary to conduct the innovation, namely (a) an extensive general description of the background, aims, and the organization of the interventions, (b) lesson scenarios describing the objectives, the necessary materials, the preferable instructional techniques, and the successive phases of each lesson, and (c) supplementary students materials, such as strategy assignment cards and reading texts. The manual was developed by the author of the article, based on extensive review of related empirical research (e.g., Bentz & Fuchs, 1996; Brown et al., 1996; Fuchs, Fuchs, Mathes, et al., 1997; Fukkink, Van der Linden, Vosse, & Vaessen, 1997; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984; Pressley et al., 1992) and prevailing manuals focusing on explicit reading comprehension instruction (e.g., Aarnoutse & van de Wouw, 1990). Teachers were not required to develop additional materials.

In addition to the manual, teachers were provided with in-service training, technical assistance and coaching by the author of the article. Prior to the implementation the underlying theoretical background of the innovations was clarified, an overview and the organization of the interventions was outlined, and the provided manual and additional materials were fully discussed during local meetings in the schools. Subsequently, during the implementation of the interventions, monthly discussions took place to exchange experiences and ideas, and to overcome practical or implementation difficulties. Moreover, two structured interviews and observations were carried out to document the fidelity of the interventions. Finally, during a last meeting shortly after the end of the intervention teachers were interviewed about their impressions and experiences with the different aspects of the innovations, and about the implementation difficulties they met. Moreover, they were invited to give their suggestions for modification and improvement of the content and organization of the interventions, the manual, and the students materials.

### *STRAT+SA and STRAT+CA Condition*

With regard to the experimental interventions both the STRAT+SA and STRAT+CA conditions are characterized by three innovative elements, namely explicit reading

strategies instruction, a sound tutor preparation, and practice of the reading strategies in weekly peer tutoring sessions.

*Explicit instruction in reading strategies.* As compared to traditional reading comprehension instruction the major substantial changes due to the explicit reading strategies instruction related to the content of teaching and learning, and to the instructional techniques. Important to mention is that both elements were identical across the experimental tutoring conditions.

In terms of the content, the innovation focused on the acquisition and mastery by the students of relevant reading strategies. With regard to the selection of strategies to focus on in the intervention, results of empirical research identifies a large variety of relevant reading strategies. On the one hand, it is impossible to address all or too many strategies in one intervention study. On the other hand, analyses of proficient readers' reading behavior revealed that skilled reading does not involve the use of a single potent strategy, but the coordination of multiple strategies while reading (Brown et al., 1996). Therefore it was necessary to make a considerate selection of relevant and feasible strategies. The selection was mainly based on an extensive review of related empirical research and innovation programs focusing on explicit strategy instruction (e.g., Brown et al., 1996; De Corte et al., 2001; Fuchs, Fuchs, Mathes, et al., 1997; Fukkink et al., 1997; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984). On the bases of these sources six strategies were selected: (a) activating prior background knowledge and linking it to the text, (b) predictive reading and verifying story outcomes, (c) distinguishing main issues from side-issues, (d) monitoring and regulating the understanding of words and expressions, (e) monitoring and regulating comprehension by tracing the ideas expressed in difficult and not understood sentences or passages, and (f) classifying types of text and adjusting reading behavior to it. Activating prior knowledge involves that students evoke what they already know about the topic of the text before they start reading and is generally considered as crucial for text comprehension (Palincsar & Brown, 1984; Paris & Oka, 1986; Pressley et al., 1992; Walraven, 1995). Within the scope of the intervention, students were more particularly instructed to infer the topic of the text on the basis of the title and accompanying illustrations, ask themselves what they already knew about that topic, and note this down in a number of key words. As regards the second reading strategy, students were taught to make predictions about upcoming events and to compare these predictions to the actual story outcomes. In this way, students are forced to look ahead while reading and to verify expectations afterwards. Distinguishing main issues from side-issues targets the skills of main idea identification and summarizing. In the present study, students more specifically learned to ask and answer who-, what-, where-, when-, and why-questions in reference to the text and to restate the main ideas of successive paragraphs of the text or of the full text. Monitoring and regulating the understanding of difficult words and expressions is essential for understanding texts as well. Students were taught to identify not understood words or expressions and to discover the meaning by looking for a definition, a synonym, or a description in the text, by using context clues, by referring to a dictionary or computer, or by enlisting someone's help. On the basis of the fifth strategy, students were encouraged to monitor and check their comprehension

and to regulate understanding of difficult sentences or passages by rereading, adjusting reading speed, or tracing the meaning of unfamiliar words or expressions. Finally, students were alerted to the fact that different types of texts - each with their own characteristics - can be distinguished. They were instructed in classifying types of text and adjusting reading behavior to it.

As concerns the instructional techniques, critical components of transactional strategies instruction (Brown et al, 1996; Pressley et al., 1992) and reciprocal teaching (Palincsar & Brown, 1984) were adopted. To begin with, much attention was paid to extensive and direct teacher explanations and modeling of strategic reasoning. This encourages students to try to use these strategies, and as they do, they construct detailed understandings of how the strategy works, the benefits it produces, and how it can be used in a variety of situations (Duffy et al., 1987; Guthrie, Cox, Knowles, Buehl, Mazzoni, & Fasulo, 2000; Pressley, 2000; Pressley & Allington, 1999). In addition, a gradual transfer from external regulation by the teacher to self-regulation of strategy use by the students was taken into account. The basic instructional model for each reading strategy consisted of the following sequence of activities. During a whole-class instructional presentation and explanation phase the teacher explicitly explained and modeled, by using the think-aloud methodology, why, how, and when a specific strategy can be helpful in enhancing comprehension. By using thinking aloud the teacher externalized as much as possible the strategic activities and processes that usually occur internally, and imparted declarative (What is the nature of the strategy?), procedural (How to deploy it?), and conditional (When to use it?) knowledge (Paris et al., 1983) of the strategies to the students. During a phase of practice and coaching using multiple examples the teacher put the strategies into practice together with the students. This phase was characterized by student assignments, systematic and explicit scaffolding and coaching by the teacher to engage students in applying and reflecting upon the strategies, and subsequent whole-class discussions. In a last phase students were encouraged to systematize the use of the strategies more independently, however with coaching by the teacher on an as-needed base. In the STRAT+CA and STRAT+SA conditions the independent practice took place in respectively cross-age and reciprocal same-age peer tutoring dyads, more extensively presented in one of the following sections.

As regards the explicit and systematic instruction in the selected reading strategies, elaborated lesson scenarios and students materials were available in the manual provided to the teachers. Student materials include strategy assignment cards and a selection of text excerpts to practice the application of the strategies. The strategy assignment cards offer structure and visual support by displaying step by step how to employ each reading strategy. The cards were used by the teachers when first modeling and explaining the strategies, and remained later as students' reference in the tutoring dyads. The texts were not especially written or adapted for the intervention, but were selected meticulously from children's literature to put the specific strategies into practice. Moreover, texts corresponding to the reading and interest level of the subjects were preferred.

*Tutor preparation.* Given the research evidence that peer tutoring is less effective when no attention is paid to a sound training of the tutors related to interpersonal and



social interaction skills (Bentz & Fuchs, 1996; Fuchs et al., 1994; Fuchs, Fuchs, Hamlett, et al., 1997), a series of lessons and materials to train students in becoming a good tutor was developed, based on related research and tutoring programs (Bentz & Fuchs, 1996; Fukkink et al., 1997). These lessons and materials were included in STRAT+CA and STRAT+SA teachers' manual and comprised instructions for role play by the teacher and the students, examples of appropriate behavior to engage in while working with a partner, worksheets for the students, and an administration card on which the tutors were expected to write down what texts they read with their tutees and which assignments had been practiced during the sessions. Verbal and visual explanations, modeling, role plays, discussions, and student practice with teacher feedback represented an important part of the tutor training.

The preparatory lessons were scheduled at the beginning of the intervention, prior to the actual strategies instruction and the peer tutoring sessions, and required seven 50-minute sessions. Tutors more particularly got acquainted with their tasks and responsibilities, they learned how to show interest, how to initiate and finish a session, how to give corrective feedback, and offer positive reinforcement for correct answers. Moreover, students were introduced to guidelines for knowing when and how to offer appropriate explanations, help, and assistance.

*Peer tutoring sessions.* Finally, both tutoring conditions were typified by peer tutoring activities, organized weekly, to systematize the use of the explicitly taught reading strategies. In the STRAT+CA condition fifth graders were paired with younger second-grade students. Children were assigned to dyads by their classroom teachers. Besides personality of the children, dyad composition was especially based on children's reading ability, so that poor and good fifth-grade readers were respectively paired with poor and good second-grade readers. In the STRAT+SA condition fifth graders were paired with classmates. Teachers assigned students to academically heterogeneous and socially compatible pairs, following the procedure described by Fuchs, Fuchs, Mathes, et al. (1997). More specifically, teachers paired all students in their class by first ranking them on reading performance and then splitting the ranked list in half. The top-ranked student in the stronger half was paired with the strongest reader in the weaker half. Next, second-ranked students in each half were paired. This matching process continued until all students had a partner. Teachers were then advised to inspect the pairings to determine whether one or more were socially incompatible. If such a dyad was found, it was changed (Fuchs, Fuchs, Mathes, et al., 1997, p. 185). In the case of an uneven number of students in the class, one group of three children was exceptionally put together, or a student from a parallel class was added to the intervention study. Taken into account the aforementioned finding that same-age partners engaging in role reciprocity made greater reading gains than partners who did not (Simmons et al., 1994), students in the same-age dyads alternated regularly between tutee and tutor roles, allowing each individual to take turns acting as the dialogue leader. Teachers saw to it that both students served as tutor for an equal amount of time. In principle dyads remained together for the duration of the school year. However, the couples were changed in the case of socially incompatible pairs. During the peer tutoring activities an important role was still reserved for the teachers. They were present and wandered about the classroom at all times to observe the

reading dyads, to coach the students and provide corrective and instructional feedback and praise, to monitor and scaffold the reading process by asking questions which force the students to evaluate the choice and use of a specific strategy, and to intervene and provide assistance or guidance on an as-needed base. Apart from the preparatory tutor training and the introductory lessons on reading strategies, peer tutoring sessions were organized once or twice a week on the average and lasted from 25 to 50 minutes, depending on the task and scheduling. Prior to each peer tutoring session a short briefing was given to the students pointing out the main ideas discussed in the previous introductory lessons. Moreover, all peer tutoring sessions were followed by a short reflection in which students' experiences and teachers' observations were discussed. During the in-service training, teachers were provided with guidelines for putting the reading dyads together, organizing the peer tutoring sessions, and coaching the reading dyads.

With regard to the succession of the introductory strategy instruction and the more independent practice in peer tutoring sessions, a "sandwich model" was applied. This implies that the teacher-led introduction of a new reading strategy was followed by at least one peer tutoring session in which the strategy was practiced when reading one of the selected text excerpts from the manual. Thereafter a number of tutoring sessions followed in which the strategy was practiced while reading books or texts the students in the reading dyads could choose themselves from the school or class library. Their choices were limited however to the range of books that matched their reading level and their specific needs. Taken into account prior research results stressing that a combination of dyadic peer interaction and structured academic activity do more to enhance cognitive gain than either of the two dimensions separately (Cohen et al., 1982; Fantuzzo, Riggio, Connelly, & Dimeff, 1989; Lambiotte et al., 1987), the aforementioned strategy assignments cards were used as a vehicle for structuring the peer tutoring interaction about the application of the reading strategies. At the end of the school year students were not only expected to make their own choice of reading books, but also to make their own choice of reading strategies and according assignment cards to work on. Therefore tutors were encouraged to take responsibility for regulating and monitoring the tutee's reading activity and to select relevant reading strategies, enhancing the chance of successful comprehension. From a theoretical perspective, consistent with Vygotsky's theory of socially mediated learning (1978), the object of the dyadic interaction in the peer tutoring activities is the joint construction of text meaning by the appropriate application of relevant reading strategies to a whole range of texts and, in the long term, the internalization and consistently self-regulative flexible use of strategic processing whenever encountering texts that are challenging to comprehend.

### *STRAT Condition*

The STRAT condition is typified by identical explicit instruction in the same six reading strategies as the STRAT+SA and STRAT+CA conditions. Teachers were not only provided with a manual including the same lesson scenarios for teaching the reading comprehension strategies and the same assignment cards and text excerpts, they also build in the gradual transfer of responsibility from teacher to students. The practice to systematize the use of the strategies did however not occur in peer tutoring

reading dyads, but was characterized by individual seatwork, teacher assistance or scaffolding on an as-needed base, whole-class teacher-led discussions, and no highly interactive peer-mediated instructional techniques. To alternate the introductory strategy lessons and the practice sessions, a sandwich model was used as well: all introduced reading strategy were followed by at least one training session applying one of the text excerpts in the manual and a number of training sessions on teacher-selected books or texts.

### *Control Condition*

The control group teachers conducted reading instruction in their typical fashion, that is without explicit instruction in reading strategies or peer tutoring. They were told that the purpose of the study was to examine primary school students' development in the field of reading comprehension achievement and were not informed that they were part of a control condition.

Due to the scope of the study we were not able to pursue a systematic and detailed analysis of reading comprehension instruction in the control classes. However, interviews with the teachers of these classes provided us with sufficient information on what went on during the time of the intervention in the experimental classes. This revealed that little or no attention was paid to the intentional and systematic teaching of reading strategies, and no peer tutoring or other forms of peer-led interaction were used. In line with previous more systematic studies of current instructional practice (e.g., Aarnoutse & Weterings, 1995; Weterings & Aarnoutse, 1986) the lessons typically involved teacher-led whole-class activities, including asking comprehension-check questions after reading a text, teacher evaluation of students' answers, and presentation of the correct answers.

Table 3.1 summarizes the main features of the interventions for each of the research conditions.

Table 3.1  
*Procedural Features of the Intervention Study for Each Research Condition*

Procedural features	Condition			
	STRAT	STRAT+SA	STRAT+CA	Control group
Pretest	✓	✓	✓	✓
Support to the teachers	✓	✓	✓	
Tutor preparation		✓	✓	
Explicit instruction in reading strategies	✓	✓	✓	
Teacher-led practice of reading strategies	✓			
Practice of reading strategies in same-age dyads		✓		
Practice of reading strategies in cross-age dyads			✓	
Posttest	✓	✓	✓	✓
Retention test	✓	✓	✓	✓

### *Time Spent on Reading Comprehension Instruction*

The interventions in the experimental conditions were not meant as an additional program on top of teachers' traditional reading comprehension classes. Experimental teachers were coached to substitute their traditional way of teaching reading

comprehension with one of the experimental programs and they were encouraged to implement the treatments during time normally allocated for reading instruction. Therefore, no differences between the four conditions were intended with regard to the total amount of time spent on reading comprehension instruction and practice.

The structured interviews with the experimental teachers yielded information on the total amount of time they spent on reading comprehension instruction on a weekly and yearly basis. The same information was obtained from all control group teachers. One-way analysis of variance on these data revealed that the different conditions allocated commensurable amounts of time to reading comprehension instruction ( $F = 0.348$ ,  $df = 3$ ,  $p = .791$ ). Post hoc analyses produced no mutual significant differences between the different conditions.

### *Measurement Instrument*

Reading comprehension achievement was assessed at pre-, post- and retention testing using the standardized test battery "Toetsen Begrijpend Lezen" (Reading Comprehension Tests) of Staphorsius and Krom (1996), which is a traditional instrument for the assessment of primary-school students' reading comprehension. At each measurement occasion a different test, increasing in level of difficulty, was used to collect the data. More specifically, the tests intended for fourth, fifth and sixth grade were used. The tests consist of three modules. Each module covers different texts, including both narrative and expository texts, followed by in total 25 multiple-choice questions. More specifically, two types of questions can be distinguished, that is questions concerning the content of a text (demanding a clear understanding of the meaning of words and sentences, the referral relation between words, the connection between sentences, and the theme of the text) and questions concerning the communication between the author and the reader of the text (e.g., objective of the author, intended target group, the author's attitude towards the matter he raises), both requiring the integration of information on different textual levels (words, sentences, successive sentences, and the text in its entirety). The three test modules vary in difficulty: the first matches the average ability of the students in the grade the test is intended for, the second and third modules are respectively more easy and more difficult. All students take the first part of the test. After discussing an example, students complete the tests individually. Next they complete the second or third module depending on their first results. In that way all students solve 50 multiple choice questions. The scores are determined by summing up the correct answers. To compare the scores of the students who finished the more easy or more difficult part of the test the sum scores are transposed into an achievement score, called the "clibscore". This clibscore ranges between 0 and 100, with the values 0, 50 and 75 showing a reading comprehension achievement respectively far below the average achievement at second grade, around the average at the end of fifth grade and above the average achievement at sixth grade. Because the standardized reading comprehension test battery is IRT-modeled, the pre-, post- and retention test clibscores are comparable. Consequently, an increase from the clibscore at pretest to the

clibscores at the post- and retention test can be interpreted as a progress in reading comprehension.

To verify the reliability of the three modules of the different tests Chronbach's  $\alpha$ -coefficients were computed for the entire sample of fifth graders, which suggest that our measures were acceptably reliable. As depicted in Table 3.2, reliability coefficients range from .66 to .81.

Table 3.2

*Chronbach's  $\alpha$ -Coefficients for the Three Modules of the Pre-, Post- and Retention Reading Comprehension Test*

Measurement occasion	Module		
	1	2	3
Pretest	.81 ( $n = 468$ )	.76 ( $n = 167$ )	.66 ( $n = 271$ )
Posttest	.72 ( $n = 442$ )	.76 ( $n = 256$ )	.74 ( $n = 175$ )
Retention test	.76 ( $n = 403$ )	.79 ( $n = 362$ )	.77 ( $n = 41$ )

### *Procedure*

The research took place in the period from September 1999 to December 2000. Attempts to teach a large number of strategies quickly, failed to produce improvements in elementary-level readers' comprehension (Paris & Oka, 1986). Therefore the implementation of the experimental interventions was spread over approximately the entire school year 1999-2000. Important to mention is that the experimental treatments were implemented by the regular class teachers with all students during time normally allocated for reading instruction. Fidelity of the implementation by the participating teachers was assured by the provided elaborated manual, the in-service training, technical assistance and coaching during monthly meetings. Structured interviews and observations in the experimental teachers' classes revealed that they all implemented the interventions in a satisfactory way and with appropriate frequency.

Reading comprehension achievement data were collected within the regular classroom context and during regularly scheduled class sessions. Pretest data collection was conducted in fifth grade at the beginning of the school year 1999-2000, prior to the initiation of the intervention program. Posttest data collection was carried out at the end of the school year after the completion of the experimental intervention. Retention test data were collected in December 2000 in sixth grade. It is to be stressed that, in accordance with the planning of the research, none of the sixth-grade teachers continued the experimental intervention.

### **Hypotheses**

The first hypothesis of the present study was that providing fifth graders with explicit reading strategies instruction would have a significant positive impact on their reading comprehension achievement. Therefore, we expected a significantly higher progress from pretest to post- and retention test on the standardized reading comprehension test

for the students in the STRAT, STRAT+SA, and STRAT+CA condition as compared to the students in the control group, characterized by traditional reading comprehension instruction. Secondly, we expected that practicing the application of the explicitly taught reading strategies during cross-age or reciprocal same-age peer tutoring activities would generate greater positive changes in reading comprehension achievement than the more traditional teacher-led practice of these strategies during whole-class activities. This hypothesis was tested by pairwise comparisons of the STRAT+SA and STRAT+CA students' progress on the one hand and the STRAT students' progress on the other hand. Finally, we hypothesized that the increase in reading comprehension achievement would be more pronounced for fifth graders functioning as tutors in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities. This prediction was tested by comparing the reading comprehension achievement growth of students in the STRAT+SA condition with that of students in the and STRAT+CA condition.

## Results

### *Data Analysis*

Because in this research 454 fifth graders are nested within a smaller number of 22 classes, it can be argued that the problem under investigation has a clear hierarchical or clustered structure. In such a sample the individual observations are generally not completely independent owing to selection processes and because of the common history and experiences individuals share by belonging to the same group (Hox, 1994). As a result, the traditional ordinary least squares regression analysis with the individual as the unit of analysis may not be used because the important assumption of independence of residual error terms is violated (Hox & Kreft, 1994). In addition, the specific research design entails that intact class groups are assigned to conditions, while observations are made on individuals within these groups. This implies that the experimental variable, assignment to one of the research conditions, is a group level variable, whereas the dependent reading comprehension variable is measured at the individual level (Krull, 1999). Moreover, cross-level interactions between explanatory variables defined at different levels of the hierarchy can influence the individual outcome variable (Hox & Kreft, 1994). Because of the joint modeling of several individual and group variables, we encounter an enduring methodological dilemma in school effectiveness research, namely the unit of analysis problem (Teddle & Reynolds, 2000). Using multilevel modeling the hierarchical nesting, the dependency, and the unit of analysis problem is handled in a natural way, for these models are specifically geared to the statistical analysis of data with a clustered structure. The application of these models results in efficient estimates of regression coefficients, correct standard errors, confidence intervals, and significance tests, which generally will be more conservative than the traditional ones which are obtained simply by ignoring the presence of clustering (Goldstein, 1995).

As multilevel models are very useful for analyzing repeated measures or longitudinal data that are collected at fixed or varying occasions (Snijders & Bosker, 1999), a

special kind of hierarchical nesting can be defined with regard to the three waves of student achievement data in the present study (pre-, post- and retention test), namely "measurement occasions nested within subjects". In this way a three-level structure arises: the level-one measurement occasions are clustered within students, who represent the level-two units and are in their turn nested within classes, which define level-three. Although it was our intention to measure all subjects at a fixed set of three measurements occasions, not each individual was attending class on all occasions. However, the occurrence of missing data - assumed they are missing at random - does not constitute a problem for the application of multilevel models in any way. An important advantage of the hierarchical linear model approach to repeated measures is the flexibility to deal with unbalanced data structures, such as repeated measures data with fixed measurement occasions where the data for some or all individuals are incomplete, or longitudinal data where some or even all individuals are measured at different sets of time points (Jones & Duncan, 1998; Snijders & Bosker, 1999).

To analyze the data, the software MLwiN for multilevel analysis was used (Rasbash et al., 1999).

### *Effect on Reading Comprehension Achievement*

To test the research hypotheses of the present study a three-step procedure was followed. The first step concerned the estimation of a three-level null model, with only an intercept term and no explanatory variables included. This model served as a baseline with which to compare subsequent more complex models and partitioned the total variance of the reading comprehension scores into three components: between-classes, between-students within-classes and between-measurements within-students variance. The second step of the analysis consisted in the input of variables at the class and student level, which possibly explain students' reading comprehension achievement scores. In particular, the measurement occasions, students' background variables, like gender and mother tongue, and the total amount of time spent on reading comprehension instruction were added to the model. Initially, these variables were included in the model as fixed effects, assuming that their impact does not vary from student to student or from class to class. Next, this assumption of a fixed linear trend was verified for each explanatory variable by allowing the coefficients to vary randomly across classes and across students within classes. Finally, the third step in the analysis concerned the inclusion of the variable "condition", which enabled us to test the hypotheses of the present study.

Further in this section we go into a full consideration of these three steps in the repeated measures multilevel analysis. Table 3.3 presents the results, using the restrictive iterative generalized least squares (RIGLS) estimation procedure.

Table 3.3

*Summary of the Model Estimates for the Three-Level Analyses of the Reading Comprehension Achievement Scores*

Parameter	Model								
	0	1	2	3a	3b	3c	3d	4a	4a (z-scores)
<i>Fixed</i>									
Intercept	41.89 (1.32)	37.66 (1.33)	37.65 (1.34)	38.34 (1.39)	38.57 (1.22)	38.67 (1.21)	38.36 (2.89)	39.79 (2.02)	-0.20 (0.13)
Post		4.24 (0.54)	4.43 (0.91)	4.42 (0.92)	4.43 (0.92)	4.28 (0.92)	5.34 (2.19)	0.92 (1.66)	0.06 (0.11)
Retention		8.98 (0.55)	8.57 (1.28)	8.56 (1.28)	8.56 (1.28)	8.37 (1.30)	6.60 (3.04)	3.33 (1.92)	0.21 (0.12)
Boy				0.40 (1.29)					
Non-native				-7.52 (3.08)	-10.96 (2.53)	-12.21 (2.76)	-10.93 (2.53)	-11.52 (2.48)	-0.74 (0.16)
Boy*non-native				-9.63 (4.96)					
Post*non-native						1.80 (1.94)			
Retention*non-native						2.47 (2.48)			
Spent time							0.01 (0.06)		
Post*spent time							-0.02 (0.05)		
Retention*spent time							0.05 (0.07)		
SA								1.99 (3.03)	0.13 (0.20)
CA								-8.05 (3.18)	-0.52 (0.20)
STRAT								-0.52 (2.60)	-0.03 (0.17)
Post*SA								3.96 (2.54)	0.25 (0.16)
Post*CA								5.60 (2.64)	0.36 (0.17)
Post* STRAT								4.84 (2.17)	0.31 (0.14)
Retention*SA								2.86 (2.95)	0.18 (0.19)
Retention*CA								11.62 (3.05)	0.75 (0.20)
Retention*STRAT								7.11 (2.52)	0.46 (0.16)
<i>Random</i>									
Level 3									
$\sigma^2_{v0}$	29.90 (11.56)	28.44 (11.08)	29.62 (11.97)	23.37 (9.99)	21.83 (9.55)	21.23 (9.35)	23.50 (10.03)	13.01 (6.81)	0.05 (0.03)
$\sigma_{v0vpost}$			-7.93 (6.16)	-7.29 (5.68)	-6.59 (5.51)	-6.01 (5.36)	-7.11 (5.77)	-3.95 (4.26)	-0.02 (0.02)
$\sigma^2_{vpost}$			13.65 (5.51)	13.85 (5.57)	13.80 (5.56)	13.39 (5.44)	14.62 (5.80)	11.30 (4.78)	0.05 (0.02)
$\sigma_{v0vret}$			-0.36 (8.09)	-1.47 (7.43)	-1.34 (7.26)	-0.94 (7.22)	-1.70 (7.56)	10.86 (5.00)	0.05 (0.02)

(table continues)



	Model								
	0	1	2	3a	3b	3c	3d	4a	4a (z-scores)
$\sigma_{\text{vpostvret}}$			4.08 (5.65)	4.21 (5.70)	4.16 (5.69)	4.13 (5.66)	4.85 (5.92)	-4.91 (3.95)	-0.02 (0.02)
$\sigma^2_{\text{vret}}$			29.01 (10.84)	29.20 (10.90)	29.14 (10.88)	29.60 (11.02)	30.21 (11.20)	14.63 (6.44)	0.06 (0.03)
Level 2									
$\sigma^2_{\mu 0}$	135.03 (11.33)	139.96 (11.14)	195.27 (13.45)	188.16 (12.96)	189.28 (13.03)	189.39 (13.04)	189.24 (13.03)	189.41 (13.04)	0.78 (0.05)
$\sigma_{\mu 0 \mu \text{post}}$			-47.40 (6.80)	-46.29 (6.66)	-46.43 (6.68)	-46.63 (6.69)	-46.37 (6.68)	-46.34 (6.68)	-0.19 (0.03)
$\sigma^2_{\mu \text{post}}$			83.98 (5.96)	83.83 (5.95)	83.88 (5.95)	84.03 (5.97)	83.83 (5.95)	83.93 (5.96)	0.35 (0.03)
$\sigma_{\mu 0 \mu \text{ret}}$			-54.23 (8.22)	-52.22 (8.05)	-52.55 (8.08)	-52.76 (8.09)	-52.61 (8.08)	-52.53 (8.07)	-0.22 (0.03)
$\sigma_{\mu \text{post} \mu \text{ret}}$			50.17 (5.75)	50.00 (5.73)	50.02 (5.74)	50.03 (5.74)	49.93 (5.73)	50.18 (5.74)	0.21 (0.02)
$\sigma^2_{\mu \text{ret}}$			119.77 (8.72)	119.57 (8.71)	119.55 (8.70)	119.65 (8.71)	119.47 (8.70)	119.67 (8.70)	0.49 (0.04)
Level 1									
$\sigma^2_{\text{e}0}$	81.26 (4.01)	61.83 (3.05)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
<i>Deviance</i>	10016	9783	9677	9655	9659	9658	9658	9628	2642
$\chi^2$		232.66	106.02	21.94	3.74	1.29	0.82	31.02	
<i>df</i>		2	10	3	2	2	3	9	
<i>p</i>		0.000	0.000	0.000	0.154	0.524	0.845	0.000	
Reference model		Model 0	Model 1	Model 2	Model 3a	Model 3b	Model 3b	Model 3b	

*Note.* Values in parentheses are standard errors.

The first step in our analysis was to examine the results of a fully unconditional three-level null model (Model 0). The intercept of 41.89 in this model is simply the overall mean of the reading comprehension clibscore across the three measurements occasions of all students in all classes. This initial analysis involves the estimation of the total variance of the dependent variable, namely 246.20, the sum of the three variance components. The total variance is further decomposed into between-classes, between-students and between-measures variance. The random part of the null model reveals that the variances at the class ( $\chi^2 = 6.69$ ,  $df = 1$ ,  $p = .010$ ), student ( $\chi^2 = 142.13$ ,  $df = 1$ ,  $p = .000$ ), and measurement occasion level ( $\chi^2 = 410.05$ ,  $df = 1$ ,  $p = .000$ ) are significantly different from zero, which provides justification for using multilevel models and indicates that each level is important: 12% of the overall variability in fifth-grade students' reading comprehension achievement scores can be attributed to class-level factors (between class differences), 55% of the variance is due to differences across individual students within the classes, and 33% is linked to differences between measurement occasions within the students. In other words, the estimates suggest that the differences in reading comprehension achievement among students within classes far outweigh the differences between classes. As expected, the differences between students are also more important than the random differences between the three measures, for in general there is typically much more variation

between individuals than between measurement occasions within individuals (Goldstein, 1995; Snijders & Bosker, 1999).

The second step in our exploration was the inclusion of explanatory variables. Since parsimonious models are preferred, only significant predictors ameliorating the model were retained. At first, to gain a clear insight into the general reading comprehension progress from pretest to respectively post- and retention test, the measurement occasions were added to the fixed part of the model (Model 1). Therefore two dummies were created with the post- and retention test contrasted against the reference group, being the pretest measure. This compound symmetry model, which is a random intercept model with no explanatory variables except for the measurement occasions (Snijders & Bosker, 1999), fits the data better than the three-level null model, for the difference in deviance of both models - which can be used as a test statistic having a chi-squared distribution, with the difference in number of parameters as degrees of freedom (Snijders & Bosker, 1999) - is highly significant ( $\chi^2 = 232.66$ ,  $df = 2$ ,  $p = .000$ ). The intercept of 37.66 is now to be considered as the overall mean pretest clibscore across all students and all classes. The fixed slope estimates on their turn show a significant progress of 4.24 ( $p = .013$ ) and 8.98 ( $p = .001$ ), respectively leading to an overall mean post- and retention test clibscore of 41.90 and 46.64. By adding the measurement occasions to the fixed part of the model, there is a major reduction in the between-measures variability. Comparing the variance components of Model 0 and 1, we see that both dummies jointly account for 24% of the unexplained level-one variance. At the same time however, the student level variance is increased by almost 4%. Although this increase in level-two variation seems typical for multilevel analysis of repeated measures, it makes it difficult to apply the suggestion by Bryk and Raudenbush (1992) to use the residual error variance of the intercept-only model as a benchmark and to examine how much this goes down by adding explanatory variables to the model (Hox, 2001).

Including the measurement occasions as predictors in the fixed part of the model assumes that the rate of change in reading comprehension from pretest to respectively post- and retention test is the same for all individuals in all classes and leads to the same coefficients for all classes and students. Nevertheless it is possible that classes, individuals within these classes or both experience a different reading comprehension growth in the course of time. To verify the compound symmetry assumption we allowed the effects of both dummies to vary randomly across all classes and students (Model 2), yielding what is called a fully multivariate model with regard to the repeated measures (Snijders & Bosker, 1999). Note that having decided that at the student level the effects of the measurement occasions could be random, it is not possible to estimate a random effect for the constant term at level one, for this coefficient is redundant in a fully multivariate model and is estimated as zero. Comparing the deviances of Model 1 and 2 reveals that the fully multivariate model is a significantly strong improvement over the compound symmetry model ( $\chi^2 = 106.02$ ,  $df = 10$ ,  $p = .000$ ). As Table 3.3 points out, the analysis shows significant variance components for the intercept and significant slope variances of both dummies at the student as well as at the class level, revealing that classes and students within those classes have different initial states and different rates of progress in the course of time.

Examining the variance function for categorical predictors, it must be clear that the variability for contrasted categories, in this case the post- and retention test, can not be read directly from the table, but equals the sum of (a) the variance of the reference category (pretest), (b) the variance of the contrast category (post- or retention test), and (c) two times the covariance of the contrast and reference category. In this respect, it can be concluded that the level-three variation between classes regarding their reading comprehension achievement is slightly decreasing from pretest (29.62) to posttest (27.40), but strongly increasing at the retention test (57.92). The level-two variation between students within classes still strongly outweigh the differences between classes and follow the same evolution with a decline from pretest (195.27) to posttest (184.44) and an increase at the retention test (206.58).

After defining the progress in reading comprehension from pretest to post- and retention test, a number of other conceivably relevant explanatory variables were added to the fixed part of the model, starting with two students' background characteristics, namely gender and mother tongue. Either of them are level-two categorical variables and represented by one dummy, that is "boy" and "non-native speaker", respectively contrasted against the categories "girl" and "native speaker of Dutch". Both main effects, as well as their interaction effect were included in the model (Model 3a), which appears to be a significant improvement over Model 2 ( $\chi^2 = 21.94$ ,  $df = 3$ ,  $p = .000$ ). However, only the main effect of being a non-native speaker is statistically significant. Therefore a more parsimonious model only including this effect was estimated (Model 3b) and found to be a just as good fit to the data, for the difference in deviance of Model 3a and 3b is not significant ( $\chi^2 = 3.74$ ,  $df = 2$ ,  $p = .154$ ). The intercept of 38.57 in Model 3b now stands for the overall mean pretest clibscore across all Dutch-speaking students across all classes. In addition, the significant fixed dummy estimate of the non-native speakers shows that children with another mother tongue than Dutch perform on average 10.96 points lower at pretest than native speakers, leading to a mean clibscore of 27.61. The underlying assumption of Model 3b is that non-native speakers progress at the same rate as Dutch-speakers, that is with 4.43 points at posttest or with 8.56 points at the retention test, respectively leading to an overall mean clibscore of 32.04 and 36.17. To test this assumption the interaction effects of the "non-native speaker" dummy with the "post-" and "retention test" dummies were added to the model (Model 3c). However, the analysis points out that Model 3c is not in any respect an improvement on Model 3b, for neither interaction effect is statistically significant, nor is the difference in deviance of both models ( $\chi^2 = 1.29$ ,  $df = 2$ ,  $p = .524$ ). This confirms the assumption of Model 3b about the similar progress of native and non-native speakers.

Next, a class-level explanatory factor was introduced into Model 3b, namely the total amount of time spent on reading comprehension instruction in fifth grade (Model 3d). Apparently, most interest goes to the interaction effects of this variable with the "post"- and "retention test" dummies, for by definition the pretest scores cannot be affected by subsequent instruction. As expected, the main effect of the included variable, which has to be considered as the influence on the pretest scores, is clearly not significant. Moreover, neither the posttest nor the retention test achievement scores are significantly affected by the total amount of time spent on reading comprehension instruction in fifth grade. The decrease in the deviance of the model is also not

statistically significant ( $\chi^2 = 0.82$ ,  $df = 3$ ,  $p = .845$ ). Therefore, a further model is built up from Model 3b.

With a view to test the postulated hypotheses the third step of the analysis consisted of adding the categorical variable "condition" to the model. To represent the four research conditions three dummy-variables were used, contrasting all experimental conditions against the control group. As we are especially interested in the differential progress of the conditions, the interaction effects with the measurement occasions were also included in the model (Model 4a). Because there is a possibility that the variance between classes differs according to the condition they are in, we allowed the effects of the three dummies representing the main effect of the conditions to vary randomly at level three. However, these more extensive models did not converge. To facilitate the interpretation of the estimates and get a better idea of the statistical power of the obtained effects, standardized regression coefficients, which can be interpreted as effect sizes, were calculated for Model 4a and included far right in Table 3.3.

Model 4a points out that by introducing the fixed main and interaction effects the deviance of the model decreases significantly ( $\chi^2 = 31.02$ ,  $df = 9$ ,  $p = .0003$ ). Besides, there is a major reduction in the variance between classes, yielding a between-classes variability of 13.01 at pretest, 16.40 at posttest and 49.36 at retention test. The intercept of 39.79 in Model 4a is now to be considered as the overall mean pretest clibscore across all Dutch-speaking students being part of the control group classes. In addition, the fixed slope estimates reveal that students in the control group make an average progress of 0.92 points at posttest and 3.33 points at retention test, respectively leading to a mean clibscore of 40.72 and 43.12 for native speakers. However, only the progress at the retention test is significantly different from zero, yet at a marginal level ( $p = .083$ ), which means that the students being part of the control group actually do not experience progress in reading comprehension in the course of the fifth grade. However, they do improve some more than one fifth of a standard deviation during the first term in sixth grade. The estimated fixed effects of the dummy variables representing the STRAT+SA, STRAT+CA and STRAT condition are respectively 1.99, -8.05, and -0.52. These effects are differential intercepts with regard to the control group, meaning that the average pretest scores of native speakers in the experimental conditions are respectively 41.78, 31.75, and 39.28. The standardized regression coefficients in Table 3.3 reveal that the STRAT+CA condition scores more than half a standard deviation lower at pretest than the control condition ( $p = .011$ ), while the other two experimental conditions do not significantly differ from this reference group. Pairwise comparisons of the experimental conditions' estimates reveal a significant lower average pretest performance of the STRAT+CA condition in comparison with the STRAT+SA ( $p = .003$ ) and STRAT condition ( $p = .011$ ) as well.

The estimates of the interaction effects between the measurement occasions and the STRAT+SA, STRAT+CA and STRAT condition are respectively 3.96 ( $p = .119$ ), 5.60 ( $p = .034$ ), and 4.84 ( $p = .025$ ) at posttest and 2.86 ( $p = .332$ ), 11.62 ( $p = .000$ ), and 7.11 ( $p = .005$ ) at retention test. These effects are differential slopes, relative to the slopes of the control condition. This implies that the average progress in clibscore for these conditions is respectively 4.88, 6.53, and 5.77 from pretest to posttest and 6.19, 14.95, and 10.43 for the whole period from pretest to retention test. The progress from

pretest to posttest of both the STRAT+CA and the STRAT condition is significantly higher than the progress of the control condition, with effect sizes of respectively 0.36 and 0.31 standard deviation. With regard to the progress over the whole period from pretest to retention test the STRAT condition experiences a significant differential improvement of 0.46 standard deviation, of which 0.31 standard deviation was already realized at posttest. Rather remarkably the STRAT+CA condition seems to affect the comprehension achievement scores equally strong in the long term, for the standardized regression coefficient reveals a differential progress from pretest to retention test of three-quarters of a standard deviation, of which almost half was gained at posttest. The improvement of the experimental STRAT+SA condition does not significantly differ from the control group, and seems to fade away from an effect size of 0.25 standard deviation at posttest to 0.18 standard deviation at retention test. Moreover, pairwise comparisons of the experimental conditions reveal that the STRAT+CA condition's progress from pretest to retention test is significantly higher than the progress of the STRAT+SA condition ( $p = .007$ ).

Table 3.4

*Model Estimates for the Three-Level Analysis of the Reading Comprehension Achievement Scores with Pre- and Retention Test Contrasted Against the Posttest as Reference*

Parameter	Model	
	4b	4b (z-scores)
<i>Fixed</i>		
Intercept	40.72 (2.14)	-0.14 (0.14)
Pretest	-0.92 (1.65)	-0.06 (0.11)
Retention	2.41 (2.66)	0.16 (0.17)
Non-native	-11.53 (2.48)	-0.74 (0.16)
SA	5.95 (3.23)	0.38 (0.21)
CA	-2.45 (3.37)	-0.16 (0.22)
STRAT	4.33 (2.77)	0.28 (0.18)
Pretest*SA	-3.96 (2.54)	-0.25 (0.16)
Pretest*CA	-5.60 (2.64)	-0.36 (0.17)
Pretest*STRAT	-4.84 (2.16)	-0.31 (0.14)
Retention*SA	-1.10 (4.15)	-0.07 (0.27)
Retention*CA	6.01 (4.22)	0.39 (0.27)
Retention*STRAT	2.26 (3.51)	0.15 (0.23)
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	16.40 (7.74)	0.07 (0.03)
$\sigma_{v0vret}$	-1.40 (7.04)	-0.01 (0.03)
$\sigma^2_{vret}$	35.77 (12.64)	0.15 (0.05)
$\sigma_{v0vpre}$	-7.36 (4.75)	-0.03 (0.02)
$\sigma_{vretvpre}$	16.21 (6.72)	0.07 (0.03)
$\sigma^2_{vpre}$	11.28 (4.77)	0.05 (0.02)
Level 2		
$\sigma^2_{\mu0}$	180.67 (12.49)	0.75 (0.05)
$\sigma_{\mu0\mu ret}$	-39.94 (7.30)	-0.17 (0.03)
$\sigma^2_{\mu ret}$	103.23 (7.59)	0.43 (0.03)
$\sigma_{\mu0\mu pre}$	-37.59 (6.42)	-0.16 (0.03)
$\sigma_{\mu ret\mu pre}$	33.74 (5.12)	0.14 (0.02)
$\sigma^2_{\mu pre}$	83.93 (5.96)	0.35 (0.03)
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)
Deviance	9628.25	2642.95

*Note.* Values in parentheses are standard errors.

To verify the statistical significance of the separate progress from posttest to retention test an additional analysis was executed with the pretest and retention test contrasted against the posttest measure as reference group (Model 4b). The results are presented in Table 3.4. The average progress from the post- to the retention test for the control group and for the STRAT+SA, STRAT+CA, and STRAT condition can be defined as respectively 2.41, 1.31, 8.42, and 4.67. Although the STRAT+CA condition experiences a differential progress of 6.01 points or 0.39 standard deviation, which is comparable to the differential progress from pretest to posttest, Table 3.4 reveals that none of the experimental conditions make a progress from posttest to retention test that is significantly higher than the progress of the control group. Pairwise comparisons of the experimental conditions' progress yields no significant differences either.

To conclude, Table 3.5 and Figure 3.1 present an overview of the predicted average reading comprehension scores for each of the four research conditions at each measurement occasion.

Table 3.5

*Predicted Average Reading Comprehension Score for Each Condition and Measurement Occasion*

Condition	Measurement occasion		
	Pretest	Posttest	Retention Test
STRAT	39.28	45.04	49.72
STRAT+SA	41.78	46.67	47.97
STRAT+CA	31.75	38.27	46.69
Control group	39.79	40.72	43.12

More particularly, the fixed estimates of Model 4a are brought together for Dutch-speaking fifth graders. Given the impact of being a non-native speaker and given the assumptions of the model about the rate of progress, an identically shaped plot could be produced for the non-native speakers, starting with a discount of 11.52 points for each research condition. The plot of the repeated measures clearly points out that the STRAT+CA students start fifth grade with a considerably lower reading comprehension performance than the other conditions. However, they start catching up in the course of fifth grade and keep improving during the first term of sixth grade. Significance tests reveal that their progress from pre- to posttest and from pretest to retention test exceeds the control condition's advancement. Moreover, the STRAT+CA condition's overall progress from pretest to retention test excels the STRAT+SA condition as well. More or less the same growth profile as the one of the STRAT+CA condition can be seen in the STRAT condition. A significance test reveals that the progress from pre- to posttest exceeds the control condition's advancement. The strong growth in the STRAT condition is however less pronounced in the period from posttest to retention test. An even more apparent decline in the second period of the research is found with regard to the STRAT+SA condition.

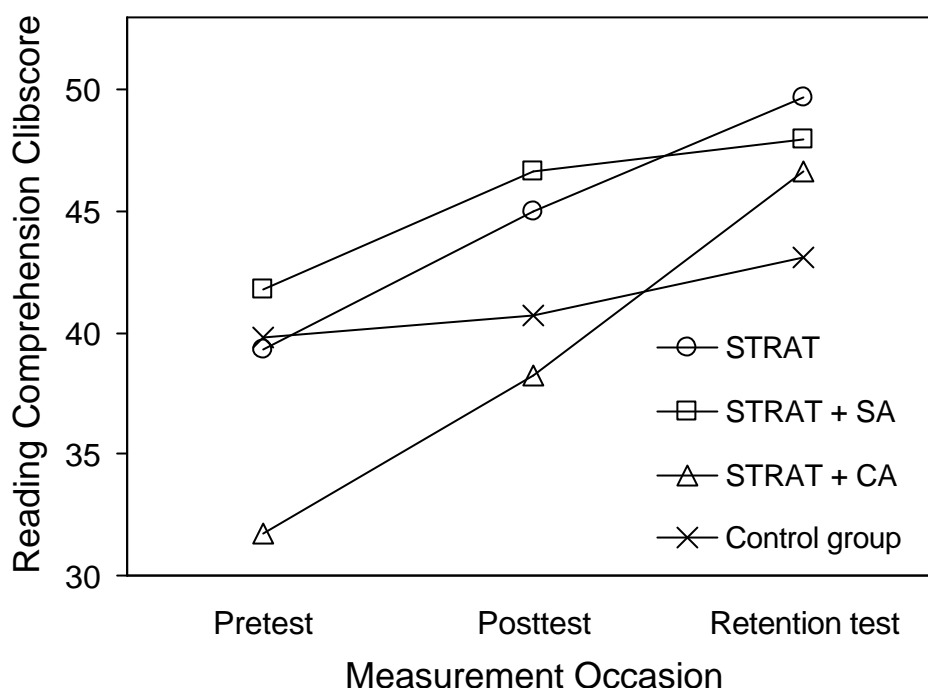


Figure 3.1

*Evolution in average reading comprehension achievement for the four research conditions from pretest to retention test.*

### *Post-Intervention Interview with the Experimental Teachers*

In addition to contributing to theory building, the present design experiment attempted to bridge the theory-practice gap. In this respect discussions and interviews with the experimental teachers and observations in their classes revealed that they all implemented the interventions in a satisfactory way. In addition, the following conclusions from the interviews shortly after completing the intervention show great promise. The teachers considered the application of the strategies as appropriate and attainable for fifth graders. Moreover, they confirmed the importance of explicit reading strategies instruction and the feasibility of implementing the interventions in substitution for their traditional way of teaching reading comprehension. With regard to the preparation of the tutors, the teachers in the experimental tutoring conditions acknowledged the necessity of a sound preparatory training. However, they suggested to intertwine the tutor training and the first steps in the practice of peer tutoring. In that way, students' enthusiasm to start can be taken into account. The innovations were not experienced as too time-consuming or overloading. Teachers appreciated the content, the instructional techniques and the organization of the interventions and were satisfied with the support and coaching before and during the implementation process. Moreover they expressed their satisfaction with regard to the elaborated manual and the student materials. Both were described as user-friendly and conveniently organized. The selection of the texts to practice the reading strategies was judged positively with respect to the content, the interest of the students, the difficulty level, and the variation in narrative and expository texts. However, teachers suggested to provide more illustrations to make the texts more attractive to the students. Finally,

teachers were very enthusiastic about their participation in the study and about their students' results. Moreover, 21 teachers were immediately willing to participate in a second research year, replicating the interventions with a new cohort students. Only one teacher dropped out because of her working half-time the following school year.

## Discussion

The main objective of the present research was to investigate the effect and durability of explicit reading strategies instruction on fifth graders' reading comprehension achievement, and to study the differential impact of practicing the strategies in cross-age and reciprocal same-age tutoring activities as compared to teacher-led practice. Therefore a pretest posttest retention test design including three experimental conditions and a commensurable control condition was used. The study was situated in the context of a relatively large number of intact classes to provide a natural setting for the implementation of the interventions by the regular class teachers. Posttest and long-term maintenance effects were studied using standardized reading comprehension tests not linked directly to the treatment and applying multilevel modeling to take the hierarchical nesting of students in classes into account.

A first hypothesis stated that explicit reading strategies instruction would enhance reading comprehension achievement scores more than traditional reading comprehension instruction, which is characterized by a strong emphasis on questioning students after reading a text and little explicit teaching of how to approach and comprehend texts. Therefore, we expected a significantly higher progress of the STRAT, STRAT+SA, and STRAT+CA condition as compared to the control group.

Generally, the research results are consistent with those obtained in previous studies and support the educational benefits of strategic reading comprehension instruction (e.g., Dole et al., 1991; Dole, 2000; Duffy et al., 1987; Haller et al., 1988, Hartman, 2001; Paris et al., 1991; Pressley, 2000; Pressley et al., 1989; Van Den Broek & Kremer, 2000). The findings more specifically corroborate the efficacy of explicit reading strategies instruction as a feasible tool to enhance fifth graders' reading comprehension achievement scores for two out of three experimental conditions. Students seem to profit most from participating in the STRAT or STRAT+CA condition. Given that teachers in the four research conditions provided similar amounts of reading instruction, it appears that the significant larger progress of both the STRAT and STRAT+CA condition can not be attributed to a greater amount of reading time for their students. Compared to the control condition both the STRAT and STRAT+CA condition made significantly larger pre- to posttest gains, with effect sizes of respectively 0.36 and 0.31 standard deviation, which compare favorably with the effect of reciprocal teaching, one of the most prominent strategy-instruction programs, developed by Palincsar and Brown (1984). For example, Rosenshine and Meister (1994) reported in their review of reciprocal teaching treatments a median effect size of 0.32 when standardized tests were used. The pretest to retention test progress even outperforms this effect, revealing effect sizes of almost half and three-quarter of a standard deviation for respectively the STRAT and STRAT+CA



condition. However, the results point out that the significant differential progress is especially situated in the period from pre- to posttest, which is exactly the phase during which the interventions were implemented and the students received the explicit reading strategies instruction. Although both significant conditions and in particular the STRAT+CA condition appear to improve more than the control condition during the first term of the sixth grade, no indication of statistical significance was detected on the measures for the post- to retention test progress. This result can be accounted for by the reported huge increase in variation between classes at retention test, leading to large standard errors. This means that classes go through a differential growth from posttest to retention test, showing differences in effectiveness of the sixth-grade teachers - who did not continued the intervention - or in underlying variables on the class level. For instructional practice the loss of significant results from post- to retention test may well imply that, in order to develop proficient strategic readers, the explicit attention to reading strategies needs to be prolonged. Clearly, further long-term research is required to ensure that the positive results of the present study are replicable and to determine the prerequisites for developing independent and self-regulative readers for life. Taken into account that the STRAT+CA condition was not equivalent on the reading comprehension measure at pretest, which is a serious threat to internal validity, replications are especially needed in reference to the STRAT+CA condition. The nonsignificant effect of the STRAT+SA condition merits particular attention as well. Particularly because the present result contrasts with previous studies focusing on the effects of reciprocal reading dyads for upper elementary-grade readers (e.g., Fuchs, Fuchs, Mathes, et al., 1997; Simmons et al., 1994). A possible explanation of this disappointing result can be found in some non-systematic observations made during the study. These indicate that the reciprocity of the tutor role in the same-age dyads faded away in the course of the intervention, as a result of which the aforementioned advantages of being a tutor were no longer evenly balanced within the dyads. Moreover, the disappointing result of the STRAT+SA condition can also be attributed to the observed decreasing amount of academic responding and less highly occupied dyadic activity and interaction during the same-age tutoring sessions as compared to the cross-age activities. With a view to test these hypotheses, more in-depth qualitatively oriented research is necessary to study the effects of the group functioning and the quality of the interaction within the peer tutoring sessions. Especially the blurred reciprocal tutor and tutee roles merit particular attention, for prior research in second through fifth grade already reported that same-age partners engaging in role reciprocity made greater reading gains than partners who did not (Simmons et al., 1994). Considering the results of a component analysis of a college-level reciprocal peer tutoring program, revealing that it is not just the mere pairing of students that is effective in increasing academic achievement, but the mutual exchange process under structured conditions (Fantuzzo et al., 1989), the declining dyadic activity and interaction in the STRAT+SA condition need to be explored as well. Taken into account that previous research reporting positive effects of same-age reciprocal reading dyads restricted the implementation of an experimental intervention to shorter periods than an entire school year, the application of several interim measures to assess short-term effects of the interventions and to explore students'

evolution throughout the school year, will also be an interesting supplement to qualitative interaction analysis in subsequent research.

With regard to the second hypothesis, it was assumed that practicing the application of the explicitly taught reading strategies during cross-age or reciprocal same-age peer tutoring activities would result in a larger progress than the more traditional teacher-led practice of these strategies during whole-class activities. This hypothesis was tested by pairwise comparisons of the progress of the STRAT+SA and STRAT+CA condition on the one hand and the STRAT condition on the other hand. Contrary to the expectations based on the literature stressing the importance of peer-led interaction on structured reading activities (e.g., Almasi, 1996; Fuchs & Fuchs, 2000; Fuchs, Fuchs, Mathes, et al., 1997; Greenwood et al., 1989; Klingner et al., 1998; Rosenshine & Meister, 1994), the progress of both tutoring conditions was not significantly different from that of the STRAT condition. Therefore, the research results did not confirm the surplus value of tutoring activities as compared to the teacher-led practice of reading strategies. However, it can be concluded that on the standardized reading comprehension test fifth graders practicing reading strategies as tutor for second-grade students, scored at least as well as the students who received teacher-led practice.

Unlike most other studies exploring the effects of peer tutoring, the present study's design allowed the explicit comparison of the effects of practicing the use of reading strategies in reciprocal same-age or in cross-age peer tutoring activities. As to the last hypothesis, it was more particularly assumed that the progress in reading comprehension would be more obvious for fifth graders functioning as tutors in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities. This prediction was tested by pairwise comparisons of the STRAT+SA and STRAT+CA condition's reading comprehension achievement growth. In the light of the results the third research hypothesis can be accepted, for there is evidence that the STRAT+CA condition's progress from pretest to retention test is significantly higher than the STRAT+SA condition's progress. The average gap between both conditions more specifically amounts to more than half a standard deviation. This result confirms the indirect evidence from the meta-analysis of Cohen et al. (1982) and suggests that the opportunity to practice the reading strategies while functioning as a tutor for a younger students plays an important part in enhancing the independent application of reading strategies and in acquiring skills in metacognitive monitoring and regulation of one's own reading process. As mentioned above, detailed in-depth observations of the quality of students' interaction and functioning in cross-age and reciprocal same-age dyads is essential to try to elucidate the differential impact of both tutoring conditions.

Apparently, the present study did not reveal pat answers to all the conceivable questions regarding the use of explicit instruction and peer tutoring activities in primary schools. Some recommendations for subsequent research were already discussed. By way of conclusion, some limitations of the study should be mentioned. Taken into account the literature on the positive effects of peer tutoring on tutors' and tutees' social and emotional functioning, a first shortcoming of the present article is the

sole focus on cognitive outcomes for only one age group. To tackle this limitation, further analyses are currently conducted to examine the benefits of explicit reading strategies instruction and social interaction about texts during class-wide cross-age or reciprocal same-age peer tutoring activities on attitudinal as well as on social and emotional outcomes, such as peer relations, and self-esteem of children of different ages.

A second comment can be made on the measurement instrument used. The fact that the present study reported significant effects on fifth graders' standardized reading comprehension tests instead of on experimenter-designed tests, ensures the strength of the experimental intervention. Examining the impact of the innovations on students' application of reading strategies and metacognitive regulation of their reading process was beyond the aims and the scope of this study. However, since explicit reading strategies instruction as well as organizing peer tutoring activities to put these strategies into practice are involved with assisting students with the process of comprehending text, tests assessing students' application of reading strategies also merit particular attention in future research. In addition, detailed in-depth analyses of the dyadic interaction and the use of thinking-aloud protocols could in this respect shed light on students' reading behavior, application of reading strategies, and metacognitive regulation of the reading process. These supplementary sources of information will enable us to ascertain whether the impact of the interventions on standardized reading comprehension tests can be attributed to an increasing mastery of reading strategies. The absence of interim measures to study the interventions' short-term effects and to explore students' evolution throughout the school year, could be viewed as a further measurement limitation of the study.

From a methodological point of view a third restriction of the present study can be mentioned. Due to the complexity of the intervention, the design does not allow to draw conclusions about the relative contribution of the different constituent components of the intervention, nor was this our purpose. In fact, it can be assumed that it is the combination of different aspects of the content and the implementation of the intervention that is responsible for the learning gains. In this respect, we speculate that the observed progress can be attributed to the focus of the intervention on the integrated teaching and acquisition of a series of different strategies, the systematic and intensive use of interactive and activating instructional techniques, the gradual transfer of responsibility in applying the strategies from teacher to students, the elaborated manual and student materials teachers could fall back on, as well as the support to the teachers before and during the interventions. However, these are mostly assumptions. A convincing explanation of what makes the STRAT and STRAT+CA condition tick, requires a different study from that we conducted: a profound component analysis can clarify the essential and perhaps dispensable components of the intervention. However, Brown et al. (1996) argue that the approach of the present research is appropriate and defensible when the interest is in evaluating the efficacy of a multicomponential instructional package.

A fourth comment can be made on the fact that the STRAT+CA students were not equivalent on reading comprehension achievement scores at pretest, which causes problems with respect to internal validity. It has to be mentioned however that the difference favors the control group and the other experimental conditions, since

weaker readers would be expected to make less progress (Juel, 1988). Taken into account that the significant lower starting position created a more stringent test of the STRAT+CA intervention, the positive research results appear to ensure the strength and educational relevance of the STRAT+CA treatment. Further research, avoiding internal validity problems, is however necessary to verify the present results. In this respect a replication study is currently performed.

Another problematic aspect of the study has to do with the absence of an in-depth analysis of what went on in the control classes. Although the available interview evidence suggested that little or no attention was paid to the intentional and systematic teaching of reading strategies and no peer tutoring or other forms of peer-led interaction were used, it was practically impossible in the framework of this study to collect systematic data about the actual instruction in all of the control classes.

Finally, critical questions can be raised about the extensive assistance and coaching of the teachers provided by the author. This intensive support to the teachers can be seen as a burden on the opportunities for implementing the innovation in instructional practice and restricts our capacity to generalize results to other situations where help of the sort we provided is absent. Given that it is unlikely that regular in-service training within the framework of implementing the innovation would have comparable levels of support available for teachers, the transfer of the present findings is uncertain. Therefore, we are currently monitoring the effects of a more restricted training of the teachers in an attempt to replicate the positive effects of the present study.

Notwithstanding the restrictions, the outcomes of the present study corroborate the efficacy of explicit reading strategies instruction as a feasible tool to enhance upper elementary-level students' reading comprehension achievement. Key findings were that (a) fifth graders benefited most from explicit reading strategies instruction supplemented with practice in teacher-led or class-wide cross-age peer tutoring activities; (b) no significant differences could be ascertained between the STRAT condition on the one hand and both tutoring conditions on the other hand; (c) the STRAT+CA condition's progress from pretest to retention test exceeds the STRAT+SA condition's improvement; (d) fifth-grade teachers confirmed the importance of explicit reading strategies instruction and the feasibility of implementing the interventions during time normally allocated to reading instruction. It is important to mention however that we do not aver that after a year of strategic reading comprehension instruction students have become self-regulated learners. Following Brown et al. (1996) our hypothesis is that "true self-regulation is the product of years of literacy experiences" (p. 34). However, the STRAT and STRAT+CA innovations in intact classes as part of the overall curriculum appear to be a promising procedure to get this process off to a good start.

## References

- Aarnoutse, C. A. J. (1991). Onderwijs in lezen [Education in reading]. In A. J. W. M. Thomassen, L. G. M. Noordman, & P. A. T. M. Eling (Eds.), *Lezen en begrijpen. De psychologie van het leesproces* (pp. 223-251). Amsterdam: Swets & Zeitlinger.

- Aarnoutse, C. A. J. (1995). *Probleemidentificatie en aanzet voor een actieplan taal* [Problem identification and initiative towards an action plan for language]. Nijmegen, The Netherlands: Vakgroep Onderwijskunde.
- Aarnoutse, C. A. J., & van de Wouw, J. (1990). *Wie dit leest. Handleiding* [Who reads this. Manual]. Tilburg, The Netherlands: Zwijzen.
- Aarnoutse, C. A. J., & Weterings, A. C. E. M. (1995). Onderwijs in begrijpend lezen [Education in reading comprehension]. *Pedagogische Studiën*, 72, 82-101.
- Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal*, 35, 309-332.
- Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.
- Almasi, J. F., McKeown, M. G., & Beck, I. L. (1996). The nature of engaged reading in classroom discussions of literature. *Journal of Literacy Research*, 28, 107-146.
- Alvermann, D. E. (2000). Classroom talk about texts: Is it dear, cheap or a bargain at any price? In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 170-192). New York: Teachers College Press.
- Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353-394). New York: Longman.
- Bargh, J. A., Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, 72, 593-604.
- Bentz, J. L., & Fuchs, L. S. (1996). Improving peers' helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly*, 19, 202-215.
- Bierman, K., & Furman, W. (1981). Effects of role assignment rationale on attitudes formed during peer tutoring. *Journal of Educational Psychology*, 73, 33-40.
- Brand-Gruwel, S., Aarnoutse, C. A. J., & Van Den Bos, K. P. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and instruction*, 8, 63-81.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432-463). New York: Macmillan.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.
- De Corte, E. (2000). Marrying theory building and the improvement of school practice: A permanent challenge for instructional psychology. *Learning and Instruction*, 10, 249-266.
- De Corte, E., Verschaffel, L., & Van De Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Educational Psychology*, 71, 531-559.

- Dole, J. A. (2000). Explicit and implicit instruction in comprehension. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 52-69). New York: Teachers College Press.
- Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61, 239-264.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliff, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putnam, J., & Bassari, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22, 347-368.
- Ehrlich, M. F., Kurtzcostes, B., & Loridant, C. (1993). Cognitive and motivational determinants of reading comprehension in good and poor readers. *Journal of Reading Behavior*, 25, 365-381.
- Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on self-concept and mathematics achievement. *Journal of Educational Psychology*, 87, 272-281.
- Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology*, 84, 331-339.
- Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *The Journal of School Psychology*, 28, 309-323.
- Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81, 173-177.
- Fitz-Gibbon, C. T. (1988). Peer tutoring as a teaching strategy. *Educational Management and Administration*, 16, 217-229.
- Fuchs, L. S., & Fuchs, D. (2000). Building student capacity to work productively during peer-assisted reading activities. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 95-115). New York: Teachers College Press.
- Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, C. L. (1994). The nature of students interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75-103.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K. & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 97, 223-249.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [Step out! Manual]. Utrecht, The Netherlands: Sardes.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

- Gourgey, A. F. (2001). Metacognition in basic skills instruction. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 33-68). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Greenwood, C. R., Carta, J. C., & Hall, R. V. (1988). The use of peer tutoring strategies in classroom management and educational instruction. *School Psychology Review*, 17, 258-275.
- Greenwood, C. R., & Delquadri, J. C. (1995). Classwide peer tutoring and the prevention of school failure. *Preventing School Failure*, 39, 21-25.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.
- Greenwood, C. R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The classwide peer tutoring program: Implementation factors moderating students achievement. *Journal of Applied Behavior Analysis*, 25, 101-116.
- Greenwood, C. R., Terry, B., Utley, C. A., Montagna, D., & Walker, D. (1993). Achievement, placement and services: Middle school benefits of classwide peer tutoring used at the elementary school. *School Psychology Review*, 22, 497-516.
- Guthrie, J. T., Cox, K. E., Knowles, K. T., Buehl, M., Mazzoni, S. A., & Fasulo, L. (2000). Building toward coherent instruction. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 209-236). New York: Teachers College Press.
- Guthrie, J. T., Schafer, W., Wang, Y. Y., Afflerbach, P. (1993). Relationships of instruction to amount of reading: An exploration of social, cognitive, and instructional connections. *Reading Research Quarterly*, 30, 8-25.
- Haller, E. P., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher*, 17, 5-8.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 33-68). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hox, J. J. (1994). *Applied multilevel analysis*. Amsterdam: TT-publikaties.
- Hox, J. J. (2001). *Multilevel analysis of regression and structural equation models*. Manuscript in preparation. Utrecht University, The Netherlands.
- Hox, J. J., & Kreft, I. (1994). Multilevel analysis methods. *Sociological Methods and Research*, 22, 283-299.
- Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: verbal versus visual strategies. *Journal of Educational Psychology*, 92, 772-782.
- Jones, K., & Duncan, C. (1998). Modeling context and heterogeneity: Applying multilevel models. In E. Scarbrough, & E. Tanenbaum (Eds.), *Research strategies in the social sciences. A guide to new approaches* (pp.95-123). Oxford, United Kingdom: Oxford University Press.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437-447.
- King-Sears, M. E., & Bradley, D. F. (1995). Classwide peer tutoring. *Preventing School Failure*, 40, 29-36.

- Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, 96, 275-293.
- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, 99, 3-21.
- Krull, J. L. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 428-444.
- Lambiotte, J. G., Dansereau, D. F., O'Donnell, A. M., Young, M., D., Skaggs, L. P., Hall, R. H., & Rocklin, T. R. (1987). Manipulating cooperative scripts for teaching and learning. *Journal of Educational Psychology*, 79, 424-430.
- Lories, G., Dardenne, B., & Yzerbyt, V. Y. (1998). From social cognition to metacognition. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition. Cognitive and social dimensions* (pp. 1-15). London: Sage.
- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.
- Mathes, P. G., Torgesen, J. K., Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal*, 38, 371-410.
- McCarthy, S. J., Hoffman, J. V., Galda, L. (1999). Readers in elementary classrooms: Learning goals and instructional principles that can inform practice. In J. J. Guthrie, & D. E. Alvermann (Eds.), *Engaged reading: Processes, practices, and policy implications* (pp. 46-80). New York: Teachers College Press.
- Morrow, L. M., & Smith, J. K. (1990). The effects of group size on interactive storybook reading. *Reading Research Quarterly*, 25, 213-231.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Palincsar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. *Remedial and Special Education*, 12, 43-53.
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 1239-1252.
- Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition and motivation. *Developmental Review*, 6, 25-56.
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Vol. 2* (p. 609-640). New York: Longman.
- Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Vol. 2* (pp. 815-860). New York: Longman.
- Pressley, M. (2000). Comprehension instruction in elementary school: A quarter-century of research progress. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 32-51). New York: Teachers College Press.



- Pressley, M., & Allington, R. (1999). What should reading instructional research be the research of? *Issues in Education*, 5, 1-35.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92, 513-555.
- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal*, 90, 3-32.
- Pressley, M., Wharton-McDonald, R., Hampston, J. M., & Echevarria, M. (1998). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies of Reading*, 2, 159-191.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, 64, 479-530.
- Roswal, G. M., Mims, A. A., Evans, M. D., Smith, B., Young, M., Burch, M., Croce, R., Horvat, M. A., & Block, M. (1995). Effects of collaborative peer tutoring on urban seventh graders. *Journal of Educational Research*, 88, 275-279.
- Simmons, D. C., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice*, 9, 203-212.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced modeling*. London: Sage.
- Staphorsius, G., & Krom, R. (1996). *Toetsen Begrijpend Lezen* [Reading Comprehension Tests]. Arnhem, The Netherlands: Cito.
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. New York: Falmer.
- Terwel, J., Gillies, R. M., van den Eeden, P., & Hoek, D. (1999). *Cooperative learning processes of students: A longitudinal multilevel perspective*. Paper presented at the American Educational Research Association Conference, Montreal, Canada (19-23 April, 1999).
- Topping, K. J. (1988). *The peer tutoring handbook: Promoting cooperative learning*. London: Croom Helm.
- Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.
- Topping, K. J. (2001). *Thinking Reading Writing. A practical guide to paired learning with peers, parents and volunteers*. London: Continuum.
- Utey, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children*, 29, 69-92.

- Van Den Broek, P., & Kremer, K. E. (2000). The mind in action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 1-31). New York: Teachers College Press.
- Van Kraayenoord, C. E. & Schneider, W. E. (1999). Reading achievement, metacognition, reading self-concept and interest: A study of German students in grades 3 and 4. *European Journal of Psychology of Education*, 14, 305-324.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Walraven, M. (1995). *Instructie in leesstrategieën. Problemen met begrijpend lezen en het effect van instructie aan zwakke lezers* [Instruction in reading strategies. Difficulties with reading comprehension and the effect of instruction for poor readers]. Amsterdam: Paedologisch Instituut.
- Webb, N. M., & Farivar, S. (1999). Developing productive group interaction in middle school mathematics. In A. M. O'Donnell, & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 117-149). Mahwah, NJ: Lawrence Erlbaum Associates.
- Weisberg, R. (1988). 1980s: A change in focus of reading comprehension research: A review of reading/learning disabilities research based on an interactive model of reading. *Learning Disability Quarterly*, 11, 149-159.
- Weterings, A. C. E. M., & Aarnoutse, C. A. J. (1986). De praktijk van het onderwijs in begrijpend lezen [Instruction in reading comprehension in practice]. *Pedagogische Studiën*, 63, 387-400.

## Chapter 4

### Effects of peer tutoring on second graders' reading fluency\*

#### Abstract

The present study was designed to test the efficacy of cross-age and reciprocal same-age peer tutoring on the reading fluency development of second-grade students in regular classrooms. The interventions were implemented weekly for the entire school year. A pretest posttest retention test design including two experimental and two control conditions was used. A multilevel linear regression model was applied to analyze the data. The analysis revealed significant fluency gains for second graders in the cross-age peer tutoring condition in the period from pretest to posttest.

#### Introduction

Generally, learning to read is considered to be a critical academic skill for future learning and adequate functioning in our modern society. Therefore reading instruction occupies an important position in elementary education. Promoting successful reading and understanding text requires instruction in both basic decoding and word recognition skills and higher-order comprehension skills. In addition, the importance of fluent reading has to be emphasized. LaBerge and Samuels (1974) were the first to introduce the concept of reading fluency as an important instructional matter. More specifically, fluency can be defined as the ability to read text with accuracy and appropriate speed and depends on well-developed word recognition skills (Chard, Simmons, & Kameenui, 1998; LaBerge & Samuels, 1974; Levy & Chard, 2001; Moyer, 1982). Fluent readers have developed automaticity and are released from the mechanisms of decoding, leaving attention free to focus on comprehension and enjoy the meaning of the text (LaBerge & Samuels, 1974; Samuels, 1979, 1997). Research supports this relationship between fluent reading and improved reading comprehension (Juel, 1991; Stanovich, 1991).

There is wide agreement that in order to read accurately and with appropriate rate, children need to have the opportunity to engage in considerable amounts of active practice (Delquadri, Greenwood, Whorton, Carta, & Hall, 1986; Greenwood, Delquadri, & Hall, 1989; Simmons, Fuchs, Fuchs, Mathes, & Pate, 1995). In educational practice, however, teachers often fail to provide sufficient opportunities to engage in active reading training. For example, the observations of Greenwood et al. (1989) in regular first- through fourth-grade classes revealed that students spent on average only 5% of the time in active reading engagement. Nonetheless, researchers have explored strategies to increase opportunities to practice reading fluency.

Well-known research-based instructional procedures that has been found effective in improving elementary and older students' oral reading fluency, are repeated reading

---

\* Based on: Van Keer, H, & Verhaeghe, J. P. Effects of peer tutoring on second graders' reading fluency development. The present chapter is submitted for publication in the *Journal of Educational Research*.

techniques in which students read texts several times until fluency is achieved. Research evidence strongly supports that readers at all proficiency levels improve their fluency (Carver & Hoffman, 1981; Dowhower, 1987; Levy, Nicholls, & Kohen, 1993; McCarthey, Hoffman, & Galda, 1999; Rashotte & Torgesen, 1985; Samuels, 1979, 1997; Sindelar, Monda, & O'Shea, 1990; Weinstein & Cooke, 1992) and comprehension (Sindelar et al., 1990) after repeated readings of the same text. Moreover, both fluency and comprehension benefits appear to be transferable to unpracticed texts (Dowhower, 1987).

Research of Eldredge, Reutzel, and Hollingsworth (1996) revealed that the Shared Reading Approach, advocated by Holdaway (1981), is even more effective in promoting fluency than are solitary repeated readings. This approach relies on teacher-supported oral reading to improve students' overall growth in reading. The approach is "teacher supported" because teachers and children read orally together. Initially, the teacher does most of the reading, but as the children become more familiar with the text, they join in and share the reading. The Shared Reading Approach is a departure from the classroom practice where individual students read in front of others and receive teacher feedback regarding their reading errors (Eldredge et al., 1996).

Besides repeated reading and shared reading, studies in the field of peer tutoring demonstrate the positive impact of literacy experiences shared with peers on fluent reading. For example, the studies on Peer-Assisted Learning Strategies (PALS) (Fuchs, Fuchs, Mathes, & Simmons, 1997; Simmons, Fuchs, Fuchs, Hodge, & Mathes, 1994) examined the effects of a class-wide same-age peer tutoring method on reading comprehension and fluency. With regard to fluent reading, the studies found significant treatment effects for second through sixth graders indicating that compared to conventional instruction, PALS students grew more on fluency after 14 to 15 weeks of intervention. Effect sizes ranged from 0.22 to 0.41 standard deviations.

Peer tutoring can be defined as "people from similar social groupings who are not professional teachers helping each other to learn, and learning themselves by teaching" (Topping, 1996, p. 322) and is characterized by specific role taking: someone has the job of tutor, while the other is in a role as tutee (Topping, 1996). This definition covers a whole series of practices that employ peers as one-on-one teachers to provide individualized instruction, practice, repetition and clarification of concepts (Topping, 1988; Utley & Mortweet, 1997). With regard to the composition of the dyads, same-age tutoring can be distinguished from cross-age tutoring. In the former variant, students are paired with classmates. The specific form of same-age tutoring in which the students alternate on a regular basis between the tutor and tutee role is called reciprocal same-age tutoring (Fantuzzo, King, & Heller, 1992). The latter variant refers to older students tutoring younger students.

The purpose of the present article, is to extend the research into the effectiveness of peer tutoring in increasing elementary students' reading fluency. As Rasinski, Padak, Linek, and Sturtevant (1994) mention, the corpus of fluency research has taken place for brief durations, outside the regular classroom context, isolated from the rest of the curriculum, with special and often small groups of students. Those study characteristics tend to limit the generalization of the results to broader instructional

contexts and populations. Conversely, the present research explored the effects of tutoring integrated into the regular curriculum in a relatively large number of real regular classrooms with the participation of all students in the course of an entire school year. More specifically, cross-age and reciprocal same-age tutoring were studied as tools for ameliorating second graders' reading fluency.

Taken into account the research results mentioned above, it can be hypothesized that reading shared with peers, as put into practice in peer tutoring dyads, will have a significant positive impact on second-grade students' oral reading fluency skills. Therefore, we expect a significantly higher progress for students reading in cross-age or reciprocal same-age peer tutoring dyads as compared to students practicing reading in a more traditional teacher-led whole-class setting. Moreover, since with the same number of peer tutoring sessions the amount of time second graders can actually spend on reading will be larger in the case of cross-age tutoring, and the corrective feedback provided by fifth graders will probably be more appropriate compared to the feedback given by second-grade peers, we also expect the positive impact to be larger in the case of cross-age peer tutoring. This hypothesis will be tested by a pairwise comparison of the progress of second graders' reading fluency in cross-age tutoring dyads on the one hand and in reciprocal same-age teams on the other hand.

## **Method**

### *Design*

A pretest posttest retention test control group design was employed in the study. More specifically, the design can be referred to as quasi-experimental, for complete naturally constituted classes were included, to ensure the ecological validity of the interventions. Participating classes were assigned to two experimental tutoring conditions and two control conditions. Reading instruction in the experimental groups was characterized by both explicit instruction in reading comprehension strategies and weekly peer tutoring sessions in either cross-age dyads (STRAT+CA) or reciprocal same-age dyads (STRAT+SA). One control group was also typified by explicit reading strategies instruction, but no peer tutoring activities were organized (STRAT). Finally, the second control group was characterized by traditional reading instruction, that is without explicit strategies instruction or peer tutoring activities.

### *Participants*

Participants for the study came from 22 second-grade classes in 18 elementary schools throughout Flanders (Belgium). In total 444 students participated. Except for one class with mainly a low SES and ethnic minority population, all schools had a predominantly white, Flemish population. The majority of the children were from middle-class families. Except for one class including only girls, there was approximately an even division of gender in the participating classes: on average 53% ( $SD = 16.54$ ) of the students were boys. At the beginning of the school year students'

age ranged from 6 to 9 years, with an average of 7 years and 4 months. The majority of the students (402) were native Dutch speakers, which is the medium of instruction in Flanders. Classes are to be considered as academically heterogeneous. Class size ranged from 15 to 28 with an average of almost 21 ( $SD = 3.50$ ) students per class, which is representative for the Flemish situation.

Second-grade teachers were on average 33 years old and had 11 years of teaching experience. Four of them were men. None of the teachers had experience in either explicit reading strategies instruction or the organization of peer tutoring activities.

Participating teachers were selected from a group of 42 second-grade teachers, all willing to take part in a long-term research study. The selection was based on the geographical distribution of the schools throughout the whole of Flanders on the one hand, and on the possibility to match teachers and classes in the different conditions as closely as possible with regard to teachers' teaching experience, class size, students' age, gender distribution, and dominating mother tongue on the other hand.

Teachers were randomly assigned to the STRAT control condition or to the experimental tutoring conditions. Within the peer tutoring conditions they were allowed to opt in favor of the STRAT+SA or STRAT+CA condition according to their own preference and the readiness of a fifth-grade colleague to collaborate in the STRAT+CA activities. The control group classes without explicit strategies instruction or peer tutoring were selected to match the teachers and class groups in the other conditions.

### *Measurement Instrument*

The Dutch standardized "Eén minuut test" (One minute test) (Brus, 1969) was used as the operational measure of reading fluency, which is a combination of accuracy and decoding speed (Chard et al., 1998). The test measures a student's ability to read words in isolation. More specifically, students read words during exactly one minute out of a list of 100 unrelated words with a gradually increasing level of difficulty. To assess accuracy and fluency, each word read is scored as correct or incorrect and the final score is determined by counting the number of words read correctly during the one minute sample. Repetitions and self-corrections were counted as correctly read words. Mispronunciations, substitutions, and omissions were counted as words read incorrectly. The test was administered individually to all second graders at all three measurement occasions, following standardized procedures according to the test's manual.

### *Interventions*

#### *STRAT+CA Condition*

Reading instruction in the experimental STRAT+CA group was characterized by explicit reading strategies instruction, a sound tutor preparation, and reading practice in weekly cross-age peer tutoring sessions. The present section shortly outlines the

implemented innovative approach. A more in-depth description of the innovations can be found in previous articles (Van Keer, 2002; Van Keer & Verhaeghe, 2002) or in the elaborated manual for participating teachers (Van Keer, in press).

As to the explicit reading strategies instruction, the intervention focused on the acquisition of six reading strategies, namely activating prior background knowledge, predictive reading, distinguishing main issues from side-issues, monitoring and regulating the understanding of difficult words, monitoring and regulating comprehension by tracing the ideas expressed in difficult sentences or passages, and classifying types of text and adjusting reading behavior to it. As concerns the instructional techniques used to teach the application of the reading strategies, much attention was paid to extensive and explicit teacher explanations and modeling of strategic reasoning. More particularly, children were instructed in strategies which support comprehension; in why, where, and when to use them; as well as in how to adapt them to various situations. In addition, a gradual transfer from external regulation by the teacher to self-regulation of strategy use by the students was striven for.

With regard to the practice of the explicitly taught reading strategies, the STRAT+CA condition was characterized by weekly organized cross-age peer tutoring activities. More specifically, second graders were paired with older fifth-grade students. Children were assigned to dyads by their classroom teachers. Besides personality of the children, dyad composition was especially based on children's reading ability, so that poor and good fifth-grade readers were respectively paired with poor and good second-grade readers. In principle dyads remained together for the duration of the school year. However, the couples were changed in the case of socially incompatible pairs. Peer tutoring sessions were organized once or twice a week on the average and lasted from 25 to 50 minutes, depending on the task and scheduling. Peer tutoring activities were organized class-wide, so all students in a class were paired and worked simultaneously.

The teacher-led introduction of a new reading strategy was followed by at least one peer tutoring session in which the strategy was put into practice while reading a text excerpt selected from the teacher's manual. Thereafter a number of tutoring sessions followed in which the strategy was practiced while reading books or texts the students in the reading dyads could choose themselves from the school or class library. Their choices were limited however to the range of books that matched their reading level and specific needs. Important to mention is that during peer tutoring sessions second graders did the reading aloud, while fifth-grade tutors monitored and regulated their tutees reading process.

Taken into account that peer tutoring appears to be less effective when no attention is paid to a sound training of the tutors (e.g., Bentz & Fuchs, 1996; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997), a series of preparatory lessons and materials was developed to train fifth graders in becoming a good tutor. The preparatory lessons were scheduled at the beginning of the intervention and required seven 50-minute sessions. Fifth graders more particularly got

acquainted with their tasks and responsibilities as tutors, they learned how to show interest, how to initiate and finish a session, how to give corrective feedback, and offer positive reinforcement for correct answers. Moreover, students were introduced to guidelines for knowing when and how to offer appropriate explanations, help, and assistance.

### *STRAT+SA Condition*

In the experimental STRAT+SA condition reading instruction was characterized by identical explicit instruction in the same six reading strategies. Contrary to the STRAT+CA condition, the strategies were however practiced in reciprocal same-age peer tutoring activities. This implies that second graders were paired with classmates. Teachers assigned students to academically heterogeneous and socially compatible pairs, following the procedure described by Fuchs, Fuchs, Mathes, et al. (1997). More specifically, teachers paired all students in their class by first ranking them on reading performance and then splitting the ranked list in half. The top-ranked student in the stronger half was paired with the strongest reader in the weaker half. Next, second-ranked students in each half were paired. This matching process continued until all students had a partner. Teachers were then advised to inspect the pairings to determine whether one or more were socially incompatible. If such a dyad was found, it was changed. In the case of an uneven number of students in the class, one group of three children was exceptionally put together, or a student from a parallel class was added to the intervention study. Taken into account that same-age partners engaging in role reciprocity made greater reading gains than partners who did not (Simmons et al., 1994), students in the same-age dyads alternated regularly between tutee and tutor roles, allowing each individual to take turns acting as the dialogue leader. This also implies that both students alternately did the reading aloud. Teachers saw to it that both students served as tutor for an equal amount of time. In principle, dyads remained together for the duration of the school year. However, the couples were changed in the case of socially incompatible pairs.

Contrary to the differences in dyad composition, the content, frequency, and organization of the cross-age and same-age peer tutoring activities ran entirely analogously. Just as fifth graders in the STRAT+CA condition all second graders in the STRAT+SA condition received a preparatory training in how to become a good tutor. These preceding activities contained the same-elements as fifth graders' preparation.

### *STRAT Control Condition*

The STRAT control condition was also typified by identical explicit instruction in the same six reading strategies as the STRAT+SA and STRAT+CA conditions, but no practice in peer tutoring activities was organized. The practice to systematize the use of the strategies was characterized by individual seatwork, teacher-led discussions, and no highly interactive peer-mediated instructional techniques.

### *Control Condition*

The control group teachers conducted reading instruction in their typical fashion, that is without explicit instruction in reading strategies or peer tutoring.



### *Support to the Teachers*

Because the interventions were implemented by the regular classroom teachers, teachers were prepared and supported to ensure that the innovations were administered with fidelity. Teachers were provided with an elaborated manual (Van Keer, in press) including all instructions, specifications, lesson scenarios, and student materials necessary to conduct the innovations. In addition, teachers were provided with regular in-service training and technical assistance by the author of the article.

To ensure that tutoring procedures were administered as planned, fidelity checks were conducted on a monthly basis. These checks were conducted by the author of the article visiting and observing participating classes during preparatory teacher-led activities and tutoring sessions.

### *Time Spent on Reading*

The interventions were not meant as an additional program on top of teachers' traditional reading classes. Teachers were encouraged to implement the treatments during time normally allocated for reading instruction. Therefore no differences between the four conditions were intended with regard to the total amount of time spent on reading instruction and practice. One-way analyses of variance on these data confirmed that the different conditions allotted commensurable amounts of time ( $F = 1.188$ ,  $df = 3$ ,  $p = .342$ ). Post hoc analyses neither produced mutual significant differences between the conditions.

### *Procedure*

The study ran from September 1999 to December 2000. The implementation of the experimental treatments was spread out over the entire school year 1999-2000 and was conducted with all students during time regularly allocated to reading instruction. The classroom teachers administered the interventions to the experimental groups.

Data were collected by the author of the article within the regular classroom context and during regularly scheduled class sessions. All four conditions were measured at three points in time. Students were pretested in October second grade before the initiation of the intervention program. Posttesting occurred during the period of May and June second grade after the completion of the experimental intervention. Finally, retention testing took place in December third grade, six months after the end of the interventions. It is to be stressed that, in accordance with the planning of the research, none of the third-grade teachers pursued the experimental intervention. Consequently, between post- and retention test (i.e., in the first term of third grade) all participants got traditional reading instruction and no experimental interventions were organized.

### *Data Analysis*

The data collected within the framework of the present study have a hierarchical or clustered structure. More specifically, 444 second graders are nested within 22 classes. Taken into account this hierarchical data structure, the application of multilevel

models is recommended, because these techniques are specifically geared to the statistical analysis of data with a clustered structure (Bryk & Raudenbush, 1992; Goldstein, 1995; Hox & Kreft, 1994). In addition, a repeated measures design was adopted, in which the measurement occasions were considered as a distinct level within the students. Consequently, three hierarchically structured levels can be distinguished: the three measurement occasions or level-one units are clustered within students who represent the level-two units, who in turn are nested within level-three classes.

To analyze the reading fluency data, the restrictive iterative generalized least squares (RIGLS) estimation procedure of the software MLwiN (Rasbash et al., 1999) was used. More specifically, a stepwise procedure was followed to build the most appropriate model with a view to test the research hypothesis. The first step concerned the estimation of a three-level null model, without explanatory variables. This model partitioned the total variance of the dependent variables into three components: between-classes, between-students, and between-measurements variance.

Subsequently, to gain a clear insight into students' general change in reading fluency from pretest to respectively post- and retention test, two dummies (i.e., post- and retention test contrasted against the pretest measure) were created and added to the model. To verify whether classes, individuals within these classes or both experience different changes in the course of time, the effects of both dummies were allowed to vary randomly across all classes and students. Note that having decided that at the student level the effects of the measurement occasions could be random, it is not possible to estimate a random effect for the constant term at level one, for this coefficient is redundant in a fully multivariate model and is estimated as zero (Snijders & Bosker, 1999).

The following step consisted in the input of conceivably relevant explanatory variables. More specifically, students' background characteristics, like gender and mother tongue, were added as additional explanatory variables. However, no significant effect on second graders' reading fluency was detected for neither of those explanatory variables. Since parsimonious models are striven for, they were dropped from the subsequent analyses.

Finally, in the last step of the analysis the effects attributable to the experimental interventions were explored. Therefore the categorical variable "condition" was added to the model. To represent the research conditions three dummy-variables were used, contrasting the STRAT+SA, STRAT+CA, and STRAT condition against the control condition. To represent the differential progress of the conditions, the interaction effects with the measurement occasions were included in the model as well.

## Results

Table 4.1 presents the results relating to the repeated measures of second graders' reading fluency and the effects of the experimental interventions. Since in this final model the effects of measurement occasion and the STRAT, STRAT+SA, and STRAT+CA conditions are "dummied out", the intercept of 33.22 is to be considered

as the overall mean pretest reading fluency score across all students being part of the control group classes. This implies that at the start of the school year control group students read on average 33.22 words per minute correctly. In addition, the estimates of the main post- and retention test effects reveal that control group students made significant improvements over time from pretest to post- and retention test. More specifically, they made an average progress of 20.92 words from pre- to posttest ( $\chi^2 = 230.57$ ,  $df = 1$ ,  $p = .000$ ) and 27.47 words from pretest to retention test ( $\chi^2 = 381.28$ ,  $df = 1$ ,  $p = .000$ ), respectively leading to a mean reading fluency score of 54.14 and 60.69 correctly read words.

Table 4.1

*Model Estimates for the Repeated Measures Three-Level Analysis of the Second-Grade Reading Fluency Scores*

Parameter	Estimates	
	Regression coefficient	Standard error
<i>Fixed</i>		
Intercept	33.22***	1.96
Posttest	20.92***	1.38
Retention test	27.47***	1.41
STRAT+SA	-4.20	3.12
STRAT+CA	0.36	3.37
STRAT	-1.96	2.60
Posttest * STRAT+SA	3.03	2.32
Posttest * STRAT+CA	4.68*	2.38
Posttest * STRAT	0.94	1.83
Retention test * STRAT+SA	3.15	2.37
Retention test * STRAT+CA	-0.56	2.43
Retention test * STRAT	0.01	1.87
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	12.93	7.20
$\sigma_{v0vpost}$	1.02	3.63
$\sigma^2_{vpost}$	8.57*	3.63
$\sigma_{v0vret}$	-5.54	4.36
$\sigma_{vpostvret}$	4.29	3.06
$\sigma^2_{vret}$	7.47*	3.79
Level 2		
$\sigma^2_{\mu0}$	205.61***	14.53
$\sigma_{\mu0\mu post}$	-60.59***	6.34
$\sigma^2_{\mu post}$	55.99***	4.10
$\sigma_{\mu0\mu ret}$	-94.43***	8.38
$\sigma_{\mu post\mu ret}$	53.80***	4.58
$\sigma^2_{\mu ret}$	86.48***	6.43
Level 1		
$\sigma^2_{\epsilon0}$	0.00	0.00

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

The estimated effects of the dummy variables representing the STRAT+SA, STRAT+CA and STRAT condition are respectively -4.20 ( $\chi^2 = 1.80$ ,  $df = 1$ ,  $p = .179$ ), 0.36 ( $\chi^2 = 0.011$ ,  $df = 1$ ,  $p = .916$ ), and -1.96 ( $\chi^2 = 0.57$ ,  $df = 1$ ,  $p = .451$ ). These effects are differential intercepts with regard to the reference condition (i.e., the control group), entailing that the average pretest scores for these conditions are respectively

31.26, 29.02, and 33.58, which is not significantly different from the average control group pretest score. Pairwise comparisons of the conditions' estimates reveal no significant pretest differences between the four research conditions.

The estimates of the interaction effects between the measurement occasions and the STRAT+SA, STRAT+CA, and STRAT conditions are respectively 3.03 ( $\chi^2 = 1.70$ ,  $df = 1$ ,  $p = .192$ ), 4.68 ( $\chi^2 = 3.86$ ,  $df = 1$ ,  $p = .049$ ), and 0.94 ( $\chi^2 = 0.27$ ,  $df = 1$ ,  $p = .607$ ) at posttest and 3.15 ( $\chi^2 = 1.77$ ,  $df = 1$ ,  $p = .183$ ), -0.56 ( $\chi^2 = 0.05$ ,  $df = 1$ ,  $p = .818$ ), and 0.01 ( $\chi^2 = 0.00$ ,  $df = 1$ ,  $p = 1.000$ ) at retention test. These effects are differential slopes, relative to the slope of the control condition, and represent the additional number of correctly read words for these conditions. This implies that the average progress is respectively 23.95, 25.60, and 21.86 from pretest to posttest and 30.62, 26.91, and 27.48 for the whole period from pretest to retention test. This pre- to post- and retention test analysis indicates that for all groups, greatest reading fluency growth occurred from pretest to posttest.

As concerns statistical significance, results indicate a significant effect favoring the STRAT+CA condition. More specifically, the analysis revealed that the STRAT+CA condition's growth from pre- to posttest was significantly higher than the progress of the control condition. With regard to the progress over the whole period from pretest to retention test, the significant differential growth however disappeared. No other significant differences between the experimental conditions on the one hand and the control group or STRAT control group on the other hand could be detected. A pairwise comparison of both experimental conditions neither yielded a significant difference.

To assess the magnitude of the intervention effects, effect sizes for the dependent reading fluency scores were calculated and presented in Table 4.2. Generally, treatment effects favored the experimental interventions. However, as mentioned before, only for the STRAT+CA condition's growth from pre- to posttest the result was significant with an effect size of 0.37 standard deviation.

Table 4.2

*Effect Sizes for the Differential Growth in Reading Fluency Skills of the STRAT+SA, STRAT+CA, and STRAT Conditions as Compared to the Control Condition*

Parameter	Effect size
Posttest * STRAT+SA	0.18
Posttest * STRAT+CA	0.37
Posttest * STRAT	0.05
Retention test * STRAT+SA	0.15
Retention test * STRAT+CA	0.07
Retention test * STRAT	0.08

An overview of the predicted average reading fluency scores for each of the four research conditions at each measurement occasion is presented in Figure 4.1. This plot of the repeated measures shows that the experimental and control conditions did not differ substantially at the start of the school year. At the posttest however, the experimental STRAT+CA group's growth outperformed the control group's evolution from pre- to posttest. Nevertheless, the differential growth did not continue at the

retention test, so the advantage of being a student in the STRAT+CA condition ceased to exist once the intervention was stopped.

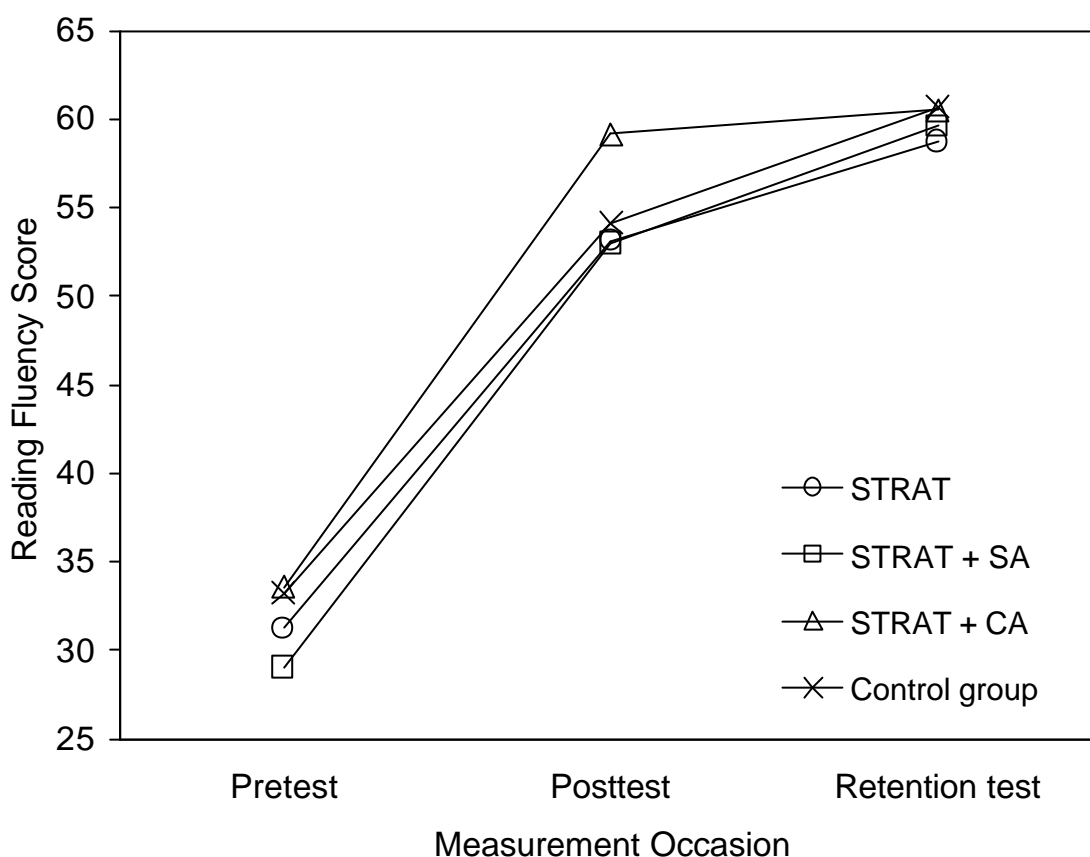


Figure 4.1

*Evolution in average reading fluency score for the four research conditions from pretest to retention test.*

As to the random part of the model, Table 4.1 reveals significant variance components for the intercept and significant slope variances with negative covariances for both post- and retention test dummies at the student level. This implies that with regard to reading fluency students had significantly different initial states and different rates of progress in the course of time, leading in fact to a decrease of the between students within classes variance from pretest (205.61) to post- (140.42) and retention test (103.24).

At the class level, a marginally significant variance component for the intercept was found, as well as significant slope variances for both measurement occasion dummies. Whereas the between students within classes variance decreased continuously, from examining the level-three variance function it can be concluded that the variation between classes first increased from pretest (12.93) to posttest (23.53) and then decreased to retention test (9.32). While being only marginally significantly different from zero at the beginning of the school year, the differences between classes regarding their average reading fluency scores became significant at the end of the school year, but diminished to a nonsignificant level within six months after the

intervention was stopped. It should be noted that the level-two variation between students within classes strongly outweighs the differences between classes.

## Discussion

The present study tested the effects of a cross-age and reciprocal same-age peer tutoring approach on the reading fluency development of second-grade students in regular classrooms. A pretest posttest retention test design including two experimental and two control conditions was used to investigate the effectiveness and the durability of the interventions. Contrary to the 14 to 15 weeks of treatment documenting the efficacy of peer tutoring on elementary-level students' reading fluency (Fuchs, Fuchs, Mathes, et al., 1997; Simmons et al., 1994), the intervention within the framework of the present study lasted an entire school year. Multilevel linear regression models were applied to analyze the data.

At first, the analysis indicated that the four research conditions did not differ significantly in their reading fluency scores at the start of the school year. Further, it was ascertained that in all conditions students made a significant progress over time. More specifically, children showed particularly considerable pre- to posttest gains. In addition, the level-two variance reflected decreasing variation between students within classes in the course of time, revealing that the reading fluency differences between students became smaller in the course of the second and third grade. The variation between classes however first increased significantly by the end of the school year, after which it decreased in the course of third grade to a level lower than the initially only marginally significant differences observed at the start of the school year in second grade.

With regard to the effectiveness of the experimental interventions, a significant treatment effect was found indicating that compared to conventional instruction in the control condition, STRAT+CA students grew more on reading fluency in the period from pretest to posttest. Aside from these significantly larger pre- to posttest gains in favor of second graders in the cross-age peer tutoring condition, no other statistically significant treatment effects were observed. Nevertheless a consistent trend favoring the reciprocal same-age condition could be noticed as well. The absence of a differential positive evolution of the STRAT control condition demonstrates that receiving explicit reading strategies instruction without extra opportunities to practice oral reading in peer tutoring dyads does not contribute to increased reading fluency skills.

These results corroborate the findings of previous research confirming the positive impact of readings shared with peers on fluent reading. Contrary to studies documenting the efficacy of same-age peer tutoring (Fuchs, Fuchs, Mathes, et al., 1997; Simmons et al., 1994), the present results however only reveal a significant effect on second graders' fluency skills for reading peer tutoring activities led by an older, fifth-grade tutor. More specifically, this effect reached an effect size of 0.37 standard deviation, which is considered as an educationally meaningful difference

(Cohen, 1988) and compares favorably with the effect sizes reported by Fuchs, Fuchs, Mathes, et al. (1997) and Simmons et al. (1994) ranging from 0.22 to 0.41 standard deviations. Although second graders' reading fluency may be graded up through cross-age tutoring, the nonsignificant STRAT+CA retention test results, six months after the end of the experimental intervention, leads us to suspect that only by pursuing the cross-age reading sessions, long-lasting treatment effects can be maintained. Future research is however necessary to verify this assumption.

With regard to the significant impact of the cross-age tutoring condition it can be assumed that peer tutoring incorporates several important principles of instruction, all making a valuable contribution to the positive effect on second graders' reading fluency. The advantage of peer-mediated methods is the creation of more favorable student-teacher ratios so the goals of individualization, frequent error correction, immediate and specific feedback, help, encouragement, reinforcement, and social support are more likely to be achieved (Delquadri et al., 1986; Fuchs, Fuchs, Mathes, et al., 1997; Greenwood & Delquadri, 1995; Utley & Mortweet, 1997). Compared to conventional forms of teacher-mediated instruction, peer tutoring programs moreover increase time on task, academic engagement, and relevant academic behaviors (Fuchs, Fuchs, Mathes, et al., 1997; Utley & Mortweet, 1997). Notwithstanding the assumption that identical benefits can be attributed to the STRAT+SA condition, students engaged in the experimental reciprocal same-age peer tutoring dyads did not make significant extra learning gains compared to students in the control group. This confirms the result of a study of Dixon-Kraus (1995), which reveals that reading with a peer of comparable ability is not likely to improve fluency as much as reading with a more able reader. However, as referred to above, the nonsignificant STRAT+SA effect contrasts with previous studies demonstrating the efficacy of same-age peer tutoring on second through sixth graders' reading fluency (Fuchs, Fuchs, Mathes, et al., 1997; Simmons et al., 1994). In this respect, however, it should be taken into account that in those earlier studies no separate analyses were done on second graders' data. The present STRAT+SA result leads us to suspect that - notwithstanding the positive trend - practicing reading in second-grade same-age pairs, as developed and administered in the present study, was too difficult for this age group to reach significant effects. Future research is necessary to explore whether the preparatory tutor training was not appropriate or unsatisfactory for second graders, whether the assignments during the peer tutoring activities were too demanding to practice independently in same-age dyads, or whether same-age peer tutoring is not an appropriate instructional technique for this second-grade age group. Another possible explanation for the present disappointing STRAT+SA result can be found in the relation between academic learning time and achievement. Within the framework of the study this relation more specifically suggests that reading achievement is positively correlated with frequent reading. More specifically, evidence indicates that teachers can not enhance student achievement simply by allocating more time within language art lessons to silent reading. Instead, teachers need to think about ways to foster diverse reading and provide scaffolds for children as they practice their reading skills, especially during the first grades of elementary school (Byrnes, 2000). In this case, the benefits of increased opportunity to read favored students in the cross-age condition who participated only as tutees led by a fifth-grade tutor and practiced reading aloud most of the time, while

students in the same-age condition alternated regularly between the tutee and tutor role.

By way of conclusion, it should be mentioned that our findings indicate a need for further research. Some recommendations for subsequent research were already discussed above. In addition, larger sample sizes at a range of elementary grade levels using a variety of measurement instruments should provide researchers with a greater understanding of the effects of peer tutoring on students' reading fluency development. More particularly, with regard to the measurement instruments, it should be taken into account that recent views about fluent reading emphasize another characteristic besides reading with accuracy and appropriate speed. According to these views, reading with fluency goes beyond accurate and rapid decoding; it also involves reading with appropriate phrasing and expression (Baker, 2000; Dowhower, 1991; Rasinski et al., 2001; Tyler & Chard, 2000). The measurement instrument used in the present study only enabled us to assess reading accuracy and rate. Therefore subsequent research should also try to measure expression and phrasing. Contrary to the utilized test, which measured student's ability to read words in isolation, attention should in this respect be called to fluency in reading context-bound texts.

Besides studying a larger range of grade levels with a greater diversity of measurement instruments, in-depth qualitatively oriented research is necessary to try to evaluate the group functioning and the quality of the interaction within the peer tutoring sessions. Presumably, qualitative interaction analysis will also enable us to account more specifically for the differential impact of cross-age and reciprocal same-age tutoring. In this respect it can be hypothesized that reading with an older student brings about more involvement and motivation and that the assistance and corrective feedback provided by fifth graders will probably be more appropriate compared to the assistance and feedback given by second-grade peers. Moreover, alternative research designs and component analyses should try to unravel the operative constituents in the peer tutoring activities.

Notwithstanding the fact that further research into the effectiveness of peer tutoring on reading fluency is called for, the results from the present study demonstrate that cross-age peer tutoring led by fifth-grade tutors may have considerable potential for improving fluent reading in second graders and may deserve a place in the regular-reading curriculum.

## References

- Baker, L. (2000). Building the word-level foundation for engaged reading. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging Young Readers. Promoting Achievement and Motivation* (pp. 17-42). New York: The Guilford Press.
- Bentz, J. L., & Fuchs, L. S. (1996). Improving peers' helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly*, 19, 202-215.
- Brus, B. Th. (1969). *Eén-minuut-test* [One minute test]. Nijmegen, The Netherlands: Berkhout.



- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Byrnes, J. P. (2000). Using instructional time effectively. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging Young Readers. Promoting Achievement and Motivation* (pp. 188-208). New York: The Guilford Press.
- Carver, R. P., & Hoffman, J.V. (1981). The effect of practice through repeated reading on gain in reading ability using a computer-based instructional system. *Reading Research Quarterly*, 16, 374-390.
- Chard, D. J., Simmons, D. C., & Kameenui, E. J. (1998). Word recognition: Research bases. In D. C. Simmons, & E. J. Kameenui (Eds.), *What Reading Research Tells Us about Children with Diverse Learning Needs. Bases and Basics* (pp. 141-167). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. New York: Academic Press.
- Delquadri, J. C., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional Children*, 52, 535-542.
- Dixon-Kraus, L. A. (1995). Partner reading and writing: Peer social dialogue and the zone of proximal development. *Journal of Reading Behavior*, 27, 45-64.
- Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly*, 22, 389-406.
- Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice*, 30, 165-173.
- Eldredge, J. L., Reutzel, D. R., & Hollingsworth, P. M. (1996). Comparing the effectiveness of two oral reading practices: Round-robin reading and the shared book experience. *Journal of Literacy Research*, 28, 201-225.
- Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology*, 84, 331-339.
- Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, C. L. (1994). The nature of students interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75-103.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K. & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 97, 223-249.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Greenwood, C. R., & Delquadri, J. C. (1995). Classwide peer tutoring and the prevention of school failure. *Preventing School Failure*, 39, 21-25.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.
- Holdaway, D. (1981). Shared book experience: Teaching reading using favorite books. *Theory into Practice*, 21, 293-300.

- Hox, J. J., & Kreft, I. (1994). Multilevel analysis methods. *Sociological Methods and Research*, 22, 283-299.
- Juel, C. (1991). Beginning reading. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. 2, pp. 759-788). New York: Longman.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 292-323.
- Levy, S., & Chard, D. J. (2001). Research on reading instruction for students with emotional and behavioural disorders. *International Journal of Disability and Education*, 48, 429-444.
- Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated readings: Process benefits for good and poor readers. *Journal of Experimental Child Psychology*, 56, 303-327.
- McCarthy, S. J., Hoffman, J. V., Galda, L. (1999). Readers in elementary classrooms: Learning goals and instructional principles that can inform practice. In J. J. Guthrie, & D. E. Alvermann (Eds.), *Engaged reading: Processes, practices, and policy implications* (pp. 46-80). New York: Teachers College Press.
- Moyer, S. B. (1982). Repeated Reading. *Journal of Learning Disabilities*, 15, 619-623.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Rashotte, C. A., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, 20, 180-202.
- Rasinski, T. V., Padak, N. D., Linek, W., & Sturtevant, E. (1994). Effects of fluency development on urban second-grade readers. *Journal of Educational Research*, 87, 158-165.
- Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher*, 32, 403-408.
- Samuels, S. J. (1997). The method of repeated readings. *The Reading Teacher*, 50, 376-381.
- Simmons, D. C., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice*, 9, 203-212.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., & Pate, H. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Sindelar, P. T., Monda, L., & O'Shea, L. (1990). Effects of repeated readings on instructional- and mastery-level readers. *Journal of Educational Research*, 83, 220-226.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced modeling*. London: Sage.
- Stanovich, K. E. (1991). Word recognition: Changing perspectives. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. 2, pp. 418-452). New York: Longman.

- Topping, K. J. (1988). *The peer tutoring handbook: Promoting cooperative learning*. London: Croom Helm.
- Topping, K. J. (1996) The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.
- Tyler, B. J., & Chard, D. J. (2000). Using readers theatre to foster fluency in struggling readers: A twist on the repeated reading strategy. *Reading and Writing Quarterly*, 16, 163-168.
- Utley, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children*, 29, 69-92.
- Van Keer, H. (in press). *Een boek voor twee. Verwerving van strategieën voor begrijpend lezen via peer tutoring*. [A book for two. Acquisition of reading comprehension strategies through peer tutoring.] Leuven, Belgium: Garant.
- Van Keer, H. (2002). *Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring*. Manuscript submitted for publication.
- Van Keer, H., & Verhaeghe, J. P. (2002). *Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions*. Manuscript submitted for publication.
- Weinstein, G., & Cooke, N. L. (1992). The effects of two repeated reading interventions on generalization of fluency. *Learning Disability Quarterly*, 15, 21-28.

## Chapter 5

### Comparing two teacher training programs for innovating reading comprehension instruction\*

#### Abstract

This study compared a year-round intensive coaching of teachers with a restricted 13-hours in-service training, designed on the basis of research-based effective teacher training components. Both trainings were developed to innovate reading comprehension instruction in elementary schools. Fourteen second- and 16 fifth-grade teachers participated. Results indicated that both conditions were equally effective in changing students' reading comprehension, reading fluency, reading strategy use, and self-efficacy perceptions. Teachers' experiences with the training were also comparable in both conditions. A significant point of difference indicated, however, that teachers attending the restricted in-service training experienced more workload with regard to settling in the innovations.

#### Introduction

Research has established convincingly that explicit instruction in the strategic aspects of processing and comprehending text and peer-led interaction about texts are more beneficial in promoting students' strategic reading and reading comprehension than traditional comprehension instruction, mainly characterized by teacher-led comprehension "testing" or questioning students about the content of a text after reading it (e.g., Almasi, 1996; Brand-Gruwel, Aarnoutse, & Van den Bos, 1998; Brown, Pressley, Van Meter, & Schuder, 1996; De Corte, Verschaffel, & Van De Ven, 2001; Dole, Duffy, Roehler, & Pearson, 1991; Duffy et al., 1987; Fuchs, Fuchs, Mathes, & Simmons, 1997; Haller, Child, & Walberg, 1988; Klingner & Vaughn, 1996; Klingner, Vaughn, & Schumm, 1998; Mathes, & Fuchs, 1994; Palincsar & Brown, 1984; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989; Pressley et al., 1992; Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995; Van Keer, 2002; Van Keer & Verhaeghe, 2002a). The successful practices emanated from intervention studies, which constructed a practical framework for effective reading comprehension instruction. Unfortunately, this was not translated into educational reality. The practice of teaching reading nowadays is still very traditional, with hardly any overt and continuous strategy instruction (Aarnoutse, 1995; Aarnoutse & Weterings, 1995; Dole, 2000; Pressley, Wharton-McDonald, Hampston, & Echevarria, 1998) or student-centered discussion (Alvermann, 2000).

Taken into account this marked gap between empirical research and instructional practice, a major issue of concern is the development and implementation of effective ways to prepare teachers to tune their teaching to recent research findings. Therefore,

---

\* Based on: Van Keer, H. & Verhaeghe, J. P. Comparing two teacher training programs for innovating reading comprehension instruction with regard to teachers' experiences and student outcomes. The present chapter is submitted for publication in *Teaching and Teacher Education*.

researchers should accept the challenge to inquire into effective strategies for disseminating research-based practices.

The focus of the present study was to compare the efficacy of two variants of teacher training as strategies to make reading comprehension instruction in elementary schools more effective. More specifically, the aforementioned research-based practices of explicit reading strategies instruction and regular student-led interactions about texts were blended in the innovative approach dealt with in the teacher training. Weekly organized peer tutoring activities were chosen as the instructional technique to create opportunities to stimulate student-centered talks about the interpretation of texts and to practice the application of the explicitly taught reading strategies.

The variants of teacher training employed in the study distinguished themselves by the amount of support lent to the teachers. Former research already confirmed the positive impact of an intensive year-round counseling, coaching, and on-the-job-training of elementary school teachers on their students' reading comprehension, reading fluency, and self-efficacy perceptions towards reading (Van Keer, 2002; Van Keer & Verhaeghe, 2002a, b). The extensive assistance provided to the teachers was, however, considered as a restriction of the study, for it is unlikely that regular in-service training authorities would be able to offer commensurable levels of support. Therefore, the aim of the present study was to develop a less intensive teacher training with effects comparable to the intensive teacher coaching.

## **Description of the Teacher Training Strategies**

This section outlines the two variants of teacher training developed to assist elementary schools in making reading comprehension instruction more effective. At first the innovative approach treated in the training will be described shortly. A more thoroughly description can be found in former publications (Van Keer, in press; Van Keer, 2002; Van Keer & Verhaeghe, 2002a). Thereafter, the organization of both teacher training variants will be presented.

### *The Innovative Approach Towards Reading Comprehension Instruction*

The discussed innovative approach towards reading comprehension instruction was identical in both variants of teacher training. More specifically, the intended instructional innovations were characterized by three components: explicit reading strategies instruction, a sound tutor preparation, and weekly organized cross-age or reciprocal same-age peer tutoring sessions.

With regard to the first component, the innovation focused on the acquisition and mastery of six reading strategies by means of modeling strategic reasoning and explicit teacher explanations of why, where, and when to use them. In addition, a gradual transfer from external regulation by the teacher to self-regulation of strategy use by the students was taken into account. Since there is evidence that peer tutoring is less effective when no attention is paid to a sound training of the tutors (Bentz & Fuchs,

1996; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997), a series of preparatory lessons and student materials focusing on the acquisition of social, communicative, and procedural skills was developed and included as the second element of the innovative approach. Finally, the innovations were typified by weekly peer tutoring activities to systematize the use of the explicitly taught reading strategies by means of structured assignments cards. Peer tutoring activities were organized class-wide, so all students in a class were paired and worked simultaneously.

### *Intensive Year-Round Coaching*

Teachers in the intensive year-round coaching condition were provided with an elaborated manual giving step by step instructions on how to conduct the instructional innovations. The manual included an extensive general description of the rationale, aims, and the organization of the innovations; lesson scenarios describing the objectives, the necessary materials, the preferable instructional techniques, and the successive phases of each lesson; and supplementary students materials, such as strategy assignment cards and reading texts. The manual was developed by the author of the article, based on extensive review of related empirical research (e.g., Bentz & Fuchs, 1996; Brown et al., 1996; Fuchs, Fuchs, Mathes, et al., 1997; Fukkink, Van der Linden, Vosse, & Vaessen, 1997; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984; Pressley et al., 1992). Teachers were not required to develop additional materials.

In addition to the manual, teachers were provided with an intensive year-round counseling, coaching, and on-the-job-training by the first author of the article. The coaching activities took place at the local schools and started with a three-hour introductory session prior to the start of the implementation of the innovations. This meeting focused on developing understanding of the characteristics of reading instruction that successfully improve students' comprehension of texts and on the general organization of the three innovative instructional components. A short video film made the innovations more concrete and outlined how to introduce them in the regular instructional practice. Moreover, room was left for questions and discussion.

After the first introduction, the lesson scenarios and additional student materials with regard to the tutor preparation and the explicit reading strategies instruction were fully discussed during six meetings. These meetings more specifically comprised an observation of a preparatory or explicit strategy instruction lesson, a detailed discussion of the attended and other already completed lessons, and a preparation of the coming lessons. Again a video film was employed to concretize the interpretation of the lesson scenarios. Prior to the first peer tutoring activity, an additional meeting was organized to discuss the composition of the reading dyads, the practical organization of the tutoring sessions, and the necessary arrangements for a smooth course. Moreover, precisely selected extracts from the video film were used to stimulate discussion and debate about the importance of the teachers' role in supervising and coaching the reading dyads.

Finally, during the implementation of the peer tutoring activities, monthly in-class observations took place, followed by discussions to exchange experiences and ideas, and to overcome practical or implementation difficulties. Generally, teachers were coached during approximately 35 hours, spread out over the entire school year.

### *Restricted 13-Hours In-Service Course*

Teachers in the restricted 13-hours in-service course condition were provided with an identical manual and the same supplementary student materials as the teachers in the coaching condition. However, the additional assistance and support offered to the teachers was less extensive, geared to the levels of support that regular in-service training authorities are able to offer. Since the aim of the present study was to develop a less intensive teacher training with effects comparable to the year-round teacher coaching, literature was searched for the essential elements in in-service courses increasing the likelihood of teacher implementation.

The predominant strategy for professional development for teachers continues to be the use of short in-service training workshops (McCutchen & Berninger, 1999; Veenman, Van Tulder, & Voeten, 1994), mainly characterized by presentation of information and discussion. Notwithstanding their popularity, however, there is considerable evidence that these one-shot teacher training strategies foster little post-training implementation of the intended changes in the classroom (Ball & Cohen, 1996; Hawley & Valli, 1999; Joyce & Showers, 1980, 1982; Leach & Ingram, 1989; Wade, 1985). Follow-up observations rarely show that teachers use the new knowledge in their daily teaching repertoires. Moreover, measures of change in student outcomes have also rarely been reported. Therefore, it can be concluded that in-service training is all too often ineffective and that the presentation of information and discussion are necessary, but not sufficient conditions to bring about change. Research, however, identified a combination of in-service training characteristics that do increase the probability of post-course teacher implementation, namely focused presentation of information and discussion about new strategies, theories or ideas to be applied, in such a format that they become acceptable and usable for the teachers; clear identification of target behaviors to be changed; demonstration and practice of relevant behaviors to be implemented; and post-course, in-class follow-up procedures, such as supervised practice and consultation, and performance feedback to the teachers (Djalil & Anderson, 1989; Harchik, Sherman, Sheldon, & Strouse, 1992; Showers, 1990; Veenman et al., 1994; Wheldall, Merrett, & Borg, 1985).

Taken into account the identified effective components of teacher training, the restricted 13-hours in-service course developed in the framework of the present study comprised an informative part, as well as in-class follow-up meetings with performance feedback to the teachers. Both components were conducted by the author of the article and were organized at the local schools. More specifically, teachers were offered three three-hours preparatory local school meetings directed at the presentation of information on the elements in the innovative approach and at the identification and

demonstration of the instructional behaviors to be implemented. In these meetings teachers were required to participate actively in the discussions and to formulate feasible plans for implementing the innovative approach in their own reading lessons. In addition, we followed the teachers back into their classes during two follow-up booster sessions, characterized by observation and consultation in the teachers' classrooms and a discussion afterwards.

The first preparatory school meeting ran completely parallel with the three-hour introductory session in the intensive coaching condition. The session more particularly focused on the presentation of information and discussion concerning the importance of explicit reading comprehension instruction, the organization of peer tutoring activities as a vehicle to encourage interaction about texts, and the significance of a sound tutor preparation. The underlying theoretical rationale and the general organization of the three innovative components were clarified and a short video film demonstrated the intended innovations. The second and third preparatory meeting dealt with the identification of the instructional behaviors to be changed by means of a discussion of the interpretation of the lesson scenarios and additional student materials, respectively with regard to the tutor preparation and the explicit reading strategies instruction. Again, the presentation of the intended instructional innovations was interspersed with demonstrations by means of video fragments. The third local meeting was completed with a discussion of the composition of the reading dyads, the practical organization of the peer tutoring activities, the necessary managerial teacher behaviors and arrangements for a smooth course, and the importance of the teachers' role in supervising and coaching the teams. Precisely selected video film extracts of reading dyads in action were used to stimulate discussion and debate and to encourage teachers to practice and express how to react and intervene to coach the teams.

Finally, the restricted in-service teacher training course was completed with two two-hours in-class follow-up sessions. These sessions started with an observation of a peer tutoring session. Thereafter, a discussion with the teachers was organized to go into the attended lesson, to give performance feedback based on the observational data, to exchange experiences and ideas, and to overcome practical or implementation difficulties.

### **Study's Purpose and Hypothesis**

As mentioned above, the aim of the present study was to compare the efficacy of two variants of teacher training as strategies to update teachers' knowledge on recent research developments in the field of reading comprehension and to translate this knowledge in classroom practice with the intention of making reading comprehension instruction in elementary schools more effective. More specifically, a year-round elaborated coaching of the teachers was equated to a more restricted 13-hours in-service course. The effectiveness of both variants of teacher training was compared by using experiences reported by the teachers, as well as student outcomes. Teachers' experiences were subdivided into satisfaction with the training course, perception of workload with regard to implementing the innovative instructional approach, and estimated student progress as a result of the implementation of the innovations. With



regard to student outcomes measures of reading comprehension, reading fluency, use of reading strategies, and self-efficacy perceptions towards reading were assessed at three different measurement occasions.

Since the restricted in-service course was composed on the basis of research-based effective components of teacher training, it was hypothesized that both teacher training strategies would be equally effective, so teachers' experiences as well as student outcomes would be comparable in both training courses. Therefore, we expected no significant differences between the intensive coaching condition and the restricted in-service course condition on the measures of teachers' reported experiences, nor on students' pre- to post- and retention test evolution with regard to reading comprehension, reading fluency, use of reading strategies, and self-efficacy perceptions towards reading.

## Method

### *Participants*

The present study is part of a larger long-term study investigating the effects of a program innovating reading comprehension instruction in Flanders. Within this framework, regular elementary school teachers were recruited two years before by an informative article and an invitation to subscribe to a teacher training program on innovative reading comprehension instruction in two widespread journals. More specifically, we focused on second- and fifth-grade teachers. Approximately 100 teachers volunteered to be engaged. Based on the geographical distribution of the schools, 14 second- and 16 fifth-grade teachers from 16 different schools throughout the whole of Flanders (Belgium) were selected to participate in the study. They were all white and had Dutch as their mother tongue. Second- and fifth-grade teachers had on average 13.25 ( $SD = 4.37$ ) and 22.29 ( $SD = 9.09$ ) years of teaching experience. Only one second-grade and five fifth-grade teachers were men.

At the time of the present research, the teachers participated for the second year in the larger long-term study. In the previous school year the teachers in the intensive coaching condition had received a similar elaborated coaching aimed at implementing the explicit reading strategies instruction, as well as the peer tutoring activities. The teachers attending the restricted teacher training course had been involved in a less elaborated in-service training directed at introducing explicit reading strategies instruction and practicing strategic reading in teacher-led whole-class activities.

In total 272 second and 342 fifth graders participated. With the exception of one inner-city school with mainly a low SES and ethnic minority population, the schools' population was comprised chiefly of white, Flemish students from middle-class families. Except for one fifth-grade class including only boys, there was approximately an even division of gender: on average 50% ( $SD = 19.19$ ) of the second- and 56% ( $SD = 20.61$ ) of the fifth-grade students were boys. The majority of the students (237 in second and 325 in fifth grade) were native Dutch speakers. Classes are to be considered as academically heterogeneous. Class size ranged from 10 to 26 with an

average of 19 ( $SD = 3.99$ ) in second grade and from 9 to 34 in fifth grade with on average 21 ( $SD = 6.88$ ) students per class, which is representative for the Flemish situation.

From the start of the long-term study onwards, participating teachers were randomly assigned to the intensive year-round coaching or restricted in-service training condition. Table 5.1 presents the number of second- and fifth-grade classes and students per condition. Univariate analysis of variance revealed no significant differences between both conditions with regard to second- and fifth-grade teachers' teaching experience ( $F = 0.47$ ;  $df = 1$ ;  $p = .508$ ;  $F = 1.84$ ;  $df = 1$ ;  $p = .200$ ), class size ( $F = 0.10$ ;  $df = 1$ ;  $p = .757$ ;  $F = 0.10$ ;  $df = 1$ ;  $p = .758$ ), gender distribution ( $F = 0.63$ ;  $df = 1$ ;  $p = .442$ ;  $F = 1.13$ ;  $df = 1$ ;  $p = .306$ ), and percentage of non-native speakers ( $F = 0.37$ ;  $df = 1$ ;  $p = .552$ ;  $F = 0.22$ ;  $df = 1$ ;  $p = .646$ ).

Table 5.1  
*Number of Participating Classes and Students*

Condition	Grade			
	2 <sup>nd</sup>		5 <sup>th</sup>	
	Classes	Students	Classes	Students
Intensive year-round coaching	7	139	7	146
Restricted 13-hours in-service training	7	133	9	196
Total	14	272	16	342

### *Instruments*

To compare the effectiveness of both teacher training conditions, teachers' experiences as well as student outcomes were assessed. Teachers' experiences were measured by means of a questionnaire concerning satisfaction with the training course, perception of workload associated with implementing the innovative instructional approach, and appraisal of student progress by the innovations. More specifically, the questionnaire was developed within the framework of the present study and incorporated 45 representative multiple-choice questions, all with a five-point Likert-type response format. With regard to student outcomes, standardized reading comprehension and fluency tests, as well as questionnaires assessing the use of reading strategies and preoccupation with attributions and self-efficacy perceptions towards reading were administered at three different measurement occasions.

This section first outlines the different parts of the teacher questionnaire. Further, the measurement instruments with regard to student outcomes will be described shortly. A more in-depth description can be found in former publications (Van Keer, 2002; Van Keer & Verhaeghe, 2002a, b).

#### *Questionnaire Concerning Teachers' Satisfaction with the Training Course*

Four items of the questionnaire assessed teachers' satisfaction with the training they attended. More specifically, they were asked to rate their overall contentment with taking part in the training course, as well as their satisfaction with the information and

counseling about the innovative program, the provided manual, and the additional student materials. The response format ranged from "strongly satisfied" (1) to "strongly dissatisfied" (5).

#### *Questionnaire Concerning Teachers' Perception of Workload*

Eleven statements examined the workload teachers experienced as a result of implementing the innovative instructional approach in their regular teaching practice. Examining the correlations between the statements showed significant overlap between certain groups of items. Therefore, responses were subjected to principal components analysis using varimax rotation. This analysis revealed three underlying factors. Five items loaded high on factor one, conceptually representing teachers' experience of workload with regard to the implementation and organization of peer tutoring activities. Three items loaded on the second factor, reflecting the workload in reference to settling in the innovative approach and preparing the students for practicing the application of reading strategies in peer tutoring dyads. For both these factors the response format of the items ranged from "experienced as a strong load" (1) to "absolutely not experienced as a load" (5). Finally, three items loaded on factor three, representing teachers' impression of increased opportunities to differentiate according to pace or content for reading dyads of different abilities as a result of the organization of peer tutoring activities. In this case, the response format of the statements ranged from "opportunities to differentiate strongly experienced " (1) to "absolutely no opportunities to differentiate experienced " (5). Factor one accounted for almost 35% of the variance and factor two explained 32% of the variance. The third factor accounted for an additional 14% of the variance. In sum the three factors explained almost 82% of the variance observed in the eleven survey items.

To verify the reliability of the three subscales Chronbach's  $\alpha$ -coefficients were computed, which revealed reliable measures for the first ( $\alpha = .90$ ;  $n = 21$ ), the second ( $\alpha = .91$ ;  $n = 19$ ), as well as the third factor ( $\alpha = .86$ ;  $n = 21$ ). Subscale scores were computed by averaging out the responses to the items within the factors. These scores were used in the subsequent analyses aimed at comparing teachers from both training conditions' experiences.

#### *Questionnaire Concerning Teachers' Appraisal of Student Progress*

Thirty statements focused on teachers' appraisal of student progress as a result of implementing the innovative instructional approach. The response format with regard to these statements ranged from "students progressed very strongly" (1) to "no progress observed" (5). More specifically, this part of the questionnaire was divided into different components, respectively assessing students' progress in the field of reading comprehension and fluency, social and interaction skills, approach to assignments, and attitude towards reading.

In four items teachers were asked to rate students' overall progress in the field of reading comprehension, as well as their advancement in the knowledge, the application, and the spontaneous use of reading strategies. In a fifth statement they evaluated students' reading fluency improvement. These five items were treated as separate dependent variables.

Fourteen items examined teachers' evaluation of student progress in the field of social and interaction skills. Principal components analysis using varimax rotation was applied to reduce these statements to a more feasible number of underlying subscales. Three factors arose, respectively accounting for almost 48%, 25%, and 9% of the variance in the original set of items. Five items loaded on factor one, representing advancement in general social and interaction skills; five items loaded on the second factor, representing evolution in social relationships and friendships. Four items loaded on the last factor, reflecting progress in the application of social and interaction skills during the peer tutoring activities. The three factors were examined for reliability. All showed acceptable internal consistency. Chronbach's  $\alpha$ -coefficients were respectively .84 ( $n = 19$ ), .88 ( $n = 18$ ), and .85 ( $n = 17$ ). Subscale scores were computed by averaging out the responses to the respective items and used in the subsequent analyses.

Eight items focused on teachers' evaluation of students' approach to assignments. More specifically, changes in responsibility, self-confidence, and autonomy were rated. Again, responses were subjected to principal components analysis using varimax rotation. This analysis revealed two underlying factors, reflecting students' approach to tasks in general and to peer tutoring assignments more specifically. Three items loaded high on factor one, five items loaded on the second factor. Factor one accounted for 51% of the variance and factor two explained 23% of the variance observed in the originally survey items. To verify the reliability of the three subscales, Chronbach's  $\alpha$ -coefficients were computed, which revealed reliable measures for the first ( $\alpha = .94$ ;  $n = 19$ ) and second factor ( $\alpha = .79$ ;  $n = 16$ ). Factor scores were computed by averaging out the responses.

Finally three items dealt with teachers' assessment of changes in students' attitude towards reading. Since the Chronbach's  $\alpha$ -coefficient of the combination of these items was highly reliable ( $\alpha = .87$ ;  $n = 22$ ), the three response scores were averaged out and used in subsequent analyses.

### *Standardized Reading Comprehension Tests*

Students' reading comprehension achievement was measured using well-established and widely used Dutch standardized test batteries. For the pretest in second grade a test from the battery "Lezen met begrip 1" (Reading with comprehension) (Verhoeven, 1993) was utilized. For the post- and retention test, two different tests from the battery "Toetsen Begrijpend Lezen" (Reading comprehension tests) (Staphorsius & Krom, 1996) were administered. The students read short expository and narrative texts and then answered multiple-choice questions with four alternatives. Scores were determined by summing up the correct answers. Cronbach's  $\alpha$ -coefficients yielded high reliability scores of .90 ( $n = 270$ ) for the pretest (30 items), .80 ( $n = 264$ ) for the posttest (25 items), and .81 ( $n = 238$ ) for the retention test (25 items).

In fifth grade reading comprehension achievement was measured with the standardized test battery "Toetsen Begrijpend Lezen" (Reading comprehension tests) (Staphorsius & Krom, 1996). The tests consisted of three modules of 25 multiple-choice questions with four alternatives each. All students took the first module of the test. Afterwards they completed the second more easy or third more difficult module, depending on their first results. Scores were determined by summing up the correct answers and

transposing the sum score into an IRT-modeled global achievement score. To verify the reliability of the three modules of the pre-, post-, and retention test, Chronbach's  $\alpha$ -coefficients were computed. Table 5.2 indicates that all reading comprehension measures are acceptably reliable. Only the most difficult module of the retention test shows a relatively low internal consistency level, which probably can be attributed to the difficulty of the test and the small number of children that completed this module.

Table 5.2

*Chronbach's  $\alpha$ -Coefficients of the Fifth-Grade Reading Comprehension Tests*

Test module	Measurement occasion <sup>a</sup>		
	Pretest	Posttest	Retention test
1	.82 ( $n = 340$ )	.75 ( $n = 316$ )	.76 ( $n = 280$ )
2	.73 ( $n = 122$ )	.83 ( $n = 177$ )	.76 ( $n = 256$ )
3	.71 ( $n = 205$ )	.74 ( $n = 130$ )	.52 ( $n = 25$ )

<sup>a</sup>At each measurement occasion a different test, increasing in level of difficulty, was used.

### *Decoding Fluency Test*

The Dutch standardized "Eén minuut test" (One minute test) (Brus, 1969) was used at each measurement occasion to assess second graders' reading fluency. More specifically, during exactly one minute individual students read words out of a list of 100 unrelated words with a gradually increasing level of difficulty. The final score was determined by counting the number of words read correctly.

### *Questionnaire on the Use of Reading Comprehension Strategies*

To estimate to what extent second- and fifth graders use reading strategies, a questionnaire was developed and administered at each measurement occasion to both second and fifth graders. More specifically, it concerns a list of 20 statements with respect to the use of strategies before, while, and after reading. Students were asked to report how often the statements applied to their own reading behavior by ticking one of the boxes ("never", "almost never", "sometimes", "very often"). Examples of the statements are: "Before I start reading, I consider what I already know about the subject of the text.", "While reading, I look for the meaning of difficult words", "While reading, I try to figure out what the story is about." Chronbach's  $\alpha$ -coefficients are presented in Table 5.3 and reveal that the questionnaire is acceptably reliable for both second and fifth graders.

### *Questionnaire on Self-Efficacy Perceptions and Related Causal Attributions*

A questionnaire was developed to measure how often students are preoccupied with positive or negative thoughts with regard to their reading ability or related causal attributions. Children were asked to report how often such thoughts crossed their mind either before, while, or after reading by ticking off "never", "almost never", "sometimes", or "very often", following statements as: "I think: I comprehended well because it was an easy text.", "I think: I did not comprehend well because there was no one to help me.", "I think: I am not good at reading". Former research (Van Keer & Verhaeghe, 2002a) revealed that success attributions and positive thoughts about one's

reading competence on the one hand and failure attributions and negative self-efficacy perceptions with regard to reading on the other hand are very closely related. Therefore, two scales were constructed. The questionnaire was administered at each measurement occasion to both second and fifth graders. Table 5.3 presents the results of the internal consistency analyses, revealing that both scales are acceptably reliable.

Table 5.3

*Chronbach's  $\alpha$ -Coefficients of the Questionnaire concerning the Use of Reading Strategies and Preoccupation with Attributions and Self-Efficacy Perceptions*

Questionnaire scale	Measurement occasion					
	Pretest		Posttest		Retention test	
	2 <sup>nd</sup> grade	5 <sup>th</sup> grade	2 <sup>nd</sup> grade	5 <sup>th</sup> grade	2 <sup>nd</sup> grade	5 <sup>th</sup> grade
Use of reading comprehension strategies	.67 ( <i>n</i> = 227)	.81 ( <i>n</i> = 280)	.77 ( <i>n</i> = 230)	.87 ( <i>n</i> = 269)	.74 ( <i>n</i> = 193)	.87 ( <i>n</i> = 248)
Success attributions and positive self-efficacy perceptions	.60 ( <i>n</i> = 247)	.68 ( <i>n</i> = 320)	.69 ( <i>n</i> = 258)	.67 ( <i>n</i> = 302)	.67 ( <i>n</i> = 225)	.71 ( <i>n</i> = 266)
Failure attributions and negative self-efficacy perceptions	.74 ( <i>n</i> = 235)	.81 ( <i>n</i> = 291)	.76 ( <i>n</i> = 239)	.86 ( <i>n</i> = 278)	.83 ( <i>n</i> = 213)	.86 ( <i>n</i> = 247)

### *Procedure*

For both conditions teacher training started in September 2000. After the first introductory training sessions, teachers started the implementation of the innovative instructional approach. All teaching activities involved in the innovative approach were conducted by the regular classroom teachers during time normally allocated for reading instruction, for one or two periods a week during the entire school year 2000-2001.

Teachers completed the questionnaire assessing their experiences at the end of the school year. Student data were collected within the regular classroom context and during regularly scheduled class sessions. Student tests were administered at three points in time: a pretest in October (second and fifth grade) before the implementation of the innovations, a posttest in May or June (second and fifth grade) after the completion of the innovations, and a retention test in December (third and sixth grade).

## **Results**

### *Experiences Reported by the Teachers*

Table 5.4 presents the average teacher response scores on the different parts of the questionnaire and their subscales. More specifically, the averages for both teacher training strategies in both grades, as well as the overall mean are displayed. To investigate whether teachers' experiences differed significantly according to grade level (second versus fifth grade) and the training they attended (intensive year-round coaching versus a restricted 13-hours in-service course), multivariate and univariate analysis of variances were executed.

Table 5.4  
*Means and Standard Deviations of the Teacher Questionnaire Results on a Five-Point Likert-Scale*

Questionnaire variables	Grade				Total
	2 <sup>nd</sup>		5 <sup>th</sup>		
	Coaching	In-service	Coaching	In-service	
<i>Satisfaction</i>					
Overall	1.80 (0.84)	1.80 (0.45)	1.83 (0.41)	1.25 (0.50)	1.70 (0.57)
Information/counseling	1.20 (0.45)	1.40 (0.55)	1.33 (0.52)	1.25 (0.50)	1.30 (0.47)
Manual	1.40 (0.89)	1.40 (0.55)	1.33 (0.52)	1.25 (0.50)	1.35 (0.59)
Student materials	2.20 (0.45)	1.80 (0.45)	2.00 (0.63)	2.75 (0.96)	2.15 (0.67)
<i>Workload</i>					
Settle in/preparation	4.42 (0.79)	3.33 (0.67)	4.44 (0.46)	3.92 (0.74)	4.12 (0.73)
Organization tutoring	3.10 (0.50)	3.47 (0.42)	3.67 (0.73)	4.10 (0.77)	3.60 (0.69)
Chance to differentiate	2.50 (0.58)	2.67 (0.33)	3.00 (0.42)	2.17 (0.43)	2.63 (0.53)
<i>Student progress</i>					
Reading comprehension	2.75 (0.50)	3.00 (0.00)	3.20 (0.45)	2.25 (0.50)	2.81 (0.54)
Strategy knowledge	3.00 (0.00)	3.00 (0.00)	2.80 (0.45)	2.00 (0.00)	2.69 (0.48)
Strategy use	2.75 (0.50)	2.67 (0.58)	3.20 (0.45)	2.25 (0.50)	2.75 (0.58)
Spontaneous strategy use	2.75 (0.50)	3.00 (0.00)	2.80 (0.45)	2.75 (0.50)	2.81 (0.40)
Reading fluency	2.17 (0.41)	2.20 (0.45)	2.83 (0.98)	2.00 (0.00)	2.35 (0.67)
General social skills	2.47 (0.42)	2.64 (0.54)	2.28 (0.33)	2.70 (0.99)	2.49 (0.49)
Social skills in tutoring	2.25 (0.66)	2.70 (0.45)	2.10 (0.49)	1.88 (0.18)	2.30 (0.54)
Social relationships	2.80 (1.06)	2.56 (0.52)	2.48 (0.58)	2.60 (0.28)	2.59 (0.59)
General task approach	2.20 (0.45)	2.67 (0.58)	2.28 (0.44)	3.00 (0.94)	2.42 (0.55)
Tutoring tasks approach	2.32 (0.52)	2.40 (0.20)	2.20 (0.36)	2.10 (0.14)	2.26 (0.36)
Attitude towards reading	2.39 (0.44)	1.94 (0.44)	2.39 (0.49)	2.25 (0.96)	2.24 (0.57)

### *Questionnaire Concerning Teachers' Satisfaction with the Training Course*

The overall mean scores in Table 5.4 indicate that teachers are generally strongly satisfied with attending the training course, especially with the information and counseling about the innovative program, and with the provided manual. The supplied student materials were somewhat less appreciated, but still reached a rather high level of contentment. A  $2 \times 2$  MANOVA was performed on the four satisfaction statements. Using Wilks' lambda criterion, the multivariate tests revealed no significant grade level ( $F(4,13) = 1.45$ ;  $p = .272$ ), condition ( $F(4,13) = 1.21$ ;  $p = .350$ ), or interaction effect ( $F(4,13) = 1.17$ ;  $p = .371$ ). As Table 5.5 reveals, the univariate tests on each of the four satisfaction items revealed no significant effects either.

### *Questionnaire Concerning Teachers' Perception of Workload*

As mentioned above, three subscales were distinguished with regard to teachers' perception of workload as a result of implementing the innovative instructional approach. The overall mean scores in Table 5.4 reveal that teachers experience little workload in reference to settling in the innovative approach and preparing the students. Little to moderate workload was experienced with regard to the implementation and organization of peer tutoring activities. In addition, teachers perceived the peer tutoring activities as a moderate to strong vehicle to build in differentiation according to pace or content for reading dyads of different abilities.

A  $2 \times 2$  MANOVA using the three subscales as the dependent measures, indicated no significant differences between the experiences of second- and fifth-grade teachers on the one hand ( $F(3,11) = 1.03$ ;  $p = .416$ ) and the intensively coached teachers and

teachers attending the restricted in-service training on the other hand ( $F(3,11) = 2.78$ ;  $p = .091$ ). No significant interaction effect of grade level and condition was found either ( $F(3,11) = 1.45$ ;  $p = .115$ ). As indicated in Table 5.5, the univariate tests, however, do reveal a significant condition effect on teachers' perceived workload in reference to settling in the innovative approach and student preparation, as well as a significant interaction effect between grade level and condition on the impression of opportunities to build in differentiation according to pace or content. More specifically, it appears that teachers attending the restricted in-service training course experience a higher workload with regard to the preparation within the framework of the innovative approach. As to the interaction effect, fifth-grade teachers attending the restricted in-service training consider the organization of the peer tutoring activities more strongly as an effective opportunity to build in and increase differentiation according to pace or content for children of different abilities.

Table 5.5

*F-values of the Univariate Tests in the 2 × 2 MANOVA's Performed on the Teacher Questionnaire Variables*

Questionnaire variables	Factors		
	Grade level	Condition	Grade level × Condition
<i>Satisfaction</i>			
Overall	1.01	1.29	1.29
Information/counseling	0.00	0.07	0.39
Manual	0.14	0.02	0.02
Student materials	1.74	0.38	4.08
<i>Workload</i>			
Settle in/preparation	0.90	6.22*	0.74
Organization tutoring	3.34	1.49	0.01
Chance to differentiate	0.00	2.17	4.88*
<i>Student progress</i>			
Reading comprehension	0.45	2.47	7.27*
Strategy knowledge	20.90**	9.29*	9.29*
Strategy use	0.00	4.18	2.94
Spontaneous strategy use	0.20	0.20	0.45
General social skills	0.05	1.07	0.19
Social skills in tutoring	3.19	0.17	1.53
Social relationships	0.15	0.03	0.25
General task approach	0.51	4.24	0.20
Tutoring tasks approach	0.97	0.00	0.18

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

### *Questionnaire Concerning Teachers' Appraisal of Student Progress*

In reference to teachers' appraisal of student progress as a result of implementing the innovative instructional approach, different components were distinguished, respectively assessing students' progress in the field of reading comprehension and fluency, social and interaction skills, approach to assignments, and attitude towards reading.

As regards the cognitive student outcomes, the overall mean scores presented in Table 5.4 reveal that teachers report a moderate to strong improvement of students' reading comprehension and fluency skills. A  $2 \times 2$  MANOVA was performed on the four reading comprehension items. The multivariate tests showed a significant grade level



( $F(4,9) = 5.99$ ;  $p = .012$ ), condition ( $F(4,9) = 6.29$ ;  $p = .011$ ), and interaction effect ( $F(4,9) = 4.83$ ;  $p = .023$ ), generally revealing that teachers in the restricted in-service training condition, especially those teaching fifth grade, report larger student advancement in the field of reading comprehension skills than the intensively coached teachers. As Table 5.5 indicates, some univariate tests reveal significant effects as well. More specifically, fifth-grade teachers and teachers in the restricted in-service training condition report a significantly larger student progress in the field of knowledge of reading comprehension strategies than second-grade teachers and intensively coached teachers. In addition, fifth-grade teachers attending the restricted 13-hours in-service training report a larger increase in general reading comprehension skills than their intensively coached counterparts.

To investigate potential grade level or condition differences with regard to teachers' appraisal of students' reading fluency progress, a two way ANOVA was performed on the reading fluency item. The analysis indicated no significant grade level ( $F = 0.62$ ;  $df = 1$ ;  $p = .442$ ) or condition main effect ( $F = 1.83$ ;  $df = 1$ ;  $p = .195$ ), nor was there a significant interaction effect ( $F = 2.14$ ;  $df = 1$ ;  $p = .162$ ).

With regard to teachers' perception of progress in the field of students' social and interaction skills, three subscales were distinguished. The overall mean scores presented in Table 5.4 indicate that teachers report a moderate to strong positive evolution in students' social relationships and friendships, as well as a moderate to strong improvement of their social and interaction skills, as a result of implementing peer tutoring activities within the framework of the innovative approach to reading comprehension instruction. The multivariate tests of a  $2 \times 2$  MANOVA on the three subscales revealed no significant grade level ( $F(3,9) = 1.84$ ;  $p = .210$ ), condition ( $F(3,9) = 2.21$ ;  $p = .156$ ), nor interaction effect ( $F(3,9) = 1.51$ ;  $p = .278$ ). As Table 5.5 reveals, univariate tests revealed no significant differences either.

In reference to teachers' perception of progress in students' approach to assignments, two distinct subscales were developed. The overall mean scores presented in Table 5.4 indicate that teachers generally report a moderate to strong positive evolution in students' responsibility, self-confidence, and autonomy when approaching peer tutoring assignments, as well as other tasks in general. A  $2 \times 2$  MANOVA on the two subscales was performed. The multivariate tests revealed no significant grade level ( $F(2,11) = 1.46$ ;  $p = .275$ ), condition ( $F(2,11) = 2.83$ ;  $p = .102$ ), nor interaction effect ( $F(2,11) = 0.38$ ;  $p = .695$ ). As indicated in Table 5.5 the univariate tests on each of the three subscales revealed no significant effects either.

Finally, to investigate grade level or condition effects on teachers' appraisal of students' progress with regard to attitude towards reading, a two-way ANOVA was performed. The analysis indicated no significant grade level ( $F = 0.38$ ;  $df = 1$ ;  $p = .546$ ) or condition main effect ( $F = 1.38$ ;  $df = 1$ ;  $p = .255$ ), nor was there a significant interaction effect ( $F = 0.38$ ;  $df = 1$ ;  $p = .546$ ).

### *Student Outcomes*

The student outcome data have a clear hierarchical structure with students nested within a smaller number of classes. More specifically, a repeated measures design was

adopted, in which the three measurement occasions were considered as a distinct level within the students. Consequently, a three-level hierarchical structure arises: the three measurement occasions (level 1) are clustered within students (level 2), in their turn nested within classes (level 3). Within this framework student data were analyzed using hierarchical linear modeling, for these models are specifically geared to take statistical account of data with clustered structure (Goldstein, 1995).

For each dependent variable, a stepwise procedure was followed to build appropriate models and to test the research hypothesis. Separate analyses were done for second and fifth graders. The parameters of the models were estimated using the restrictive iterative generalized least squares (RIGLS) estimation procedure of the software MLwiN (Rasbash et al., 1999). Since parsimonious models are preferred, only significant estimates ameliorating the model were retained.

### *Reading Comprehension Achievement*

Besides the effects for the measurement occasions, the final repeated measures model with regard to second and fifth graders' reading comprehension achievement included explanatory variables such as gender, mother tongue and the number of years students are behind at school. For second grade reading fluency scores were also included in the model. More specifically, it appears that second-grade boys perform significantly lower than girls with a fixed arrear at each measurement occasion. A similar effect was found for being a non-native Dutch speaker in fifth grade. On the average second-grade non-native speakers also perform significantly lower at pretest, but they appeared to catch up significantly in the course of second and third grade. Further, the number of years second and fifth graders are behind at school has been found inversely proportional to their comprehension scores, with a fixed impact for all the measurement occasions. Finally, as regards the impact of reading fluency it can be concluded that second graders with higher pretest fluency scores on the average have higher pretest reading comprehension scores as well. Nevertheless, the relationship between reading fluency and reading comprehension becomes weaker at the post- and retention test.

Having controlled for the aforementioned explanatory variables, the results revealed that at pretest second- ( $\chi^2 = 0.17$ ,  $df = 1$ ,  $p = .684$ ) and fifth-grade students ( $\chi^2 = 3.75$ ,  $df = 1$ ,  $p = .053$ ) with teachers in the restricted training condition did not significantly differ from their peers with teachers in the intensive coaching condition. Moreover, no significant differences were found between both research conditions with regard to students' progress in reading comprehension achievement from pretest to posttest in second ( $\chi^2 = 0.83$ ,  $df = 1$ ,  $p = .362$ ) and fifth grade ( $\chi^2 = 0.06$ ,  $df = 1$ ,  $p = .805$ ), nor from pretest to retention test in second ( $\chi^2 = 1.51$ ,  $df = 1$ ,  $p = .219$ ) and fifth grade ( $\chi^2 = 0.09$ ,  $df = 1$ ,  $p = .764$ ).

### *Reading Fluency Achievement*

Second graders were also tested with regard to their reading fluency achievement. The final multilevel model indicated that second-grade boys performed significantly lower than girls with a fixed arrear at each measurement occasion and that at pretest no significant differences arose between students with teachers in the restricted training or intensive coaching condition ( $\chi^2 = 0.17$ ,  $df = 1$ ,  $p = .685$ ). In addition, no differential

reading fluency progress was found from pretest to post- ( $\chi^2 = 1.06$ ,  $df = 1$ ,  $p = .303$ ) and retention test ( $\chi^2 = 0.17$ ,  $df = 1$ ,  $p = .683$ ).

#### *Use of Reading Comprehension Strategies*

With regard to students' report of strategy use, similar models were built up. It appeared that at each measurement occasion second-grade boys reported significantly less strategic reading than girls. No significant pretest differences between the research conditions were found, neither for second ( $\chi^2 = 0.24$ ,  $df = 1$ ,  $p = .622$ ), nor for fifth grade ( $\chi^2 = 3.29$ ,  $df = 1$ ,  $p = .070$ ). Moreover, no differential increase or decrease in the reports of reading strategy use of students from both conditions was found from pretest to posttest in second ( $\chi^2 = 2.55$ ,  $df = 1$ ,  $p = .110$ ) and fifth grade ( $\chi^2 = 2.626$ ,  $df = 1$ ,  $p = .105$ ), nor from pretest to retention test in second ( $\chi^2 = 0.00$ ,  $df = 1$ ,  $p = .964$ ) and fifth grade ( $\chi^2 = 3.53$ ,  $df = 1$ ,  $p = .060$ ).

#### *Self-Efficacy Perceptions and Related Causal Attributions*

The final multilevel models with regard to students' self-efficacy measures reveal that besides the dummies representing the measurement occasions second graders' gender and fifth graders' mother tongue yielded significant estimates and improvements of the models. More specifically, at each measurement occasion second-grade boys reported significantly less thoughts with regard to success attributions and positive self-efficacy perceptions than girls. As concerns the fifth graders, at each measurement occasion non-native students reported significantly more thoughts with regard to failure attributions and negative self-efficacy perceptions than Dutch-speaking students.

Having controlled for the aforementioned explanatory variables, the models with regard to second and fifth graders' thoughts relating to success attributions and positive self-efficacy perceptions towards reading indicate no significant pretest differences between both research conditions, neither for second ( $\chi^2 = 0.00$ ,  $df = 1$ ,  $p = .964$ ), nor for fifth grade ( $\chi^2 = 0.69$ ,  $df = 1$ ,  $p = .407$ ). Moreover, no significantly differential evolution was found from pretest to posttest in second ( $\chi^2 = 0.16$ ,  $df = 1$ ,  $p = .686$ ) and fifth grade ( $\chi^2 = 0.09$ ,  $df = 1$ ,  $p = .762$ ), nor from pretest to retention test in second ( $\chi^2 = 1.54$ ,  $df = 1$ ,  $p = .214$ ) and fifth grade ( $\chi^2 = 0.82$ ,  $df = 1$ ,  $p = .365$ ).

The models with regard to second and fifth graders' thoughts relating to failure attributions and negative self-efficacy perceptions indicate similar results: no significant pretest differences between the second- ( $\chi^2 = 3.30$ ,  $df = 1$ ,  $p = .069$ ) and fifth-grade ( $\chi^2 = 1.27$ ,  $df = 1$ ,  $p = .260$ ) research conditions; and no significantly differential evolution from pretest to posttest in second ( $\chi^2 = 0.00$ ,  $df = 1$ ,  $p = .950$ ) and fifth grade ( $\chi^2 = 2.83$ ,  $df = 1$ ,  $p = .093$ ), nor from pretest to retention test in second ( $\chi^2 = 1.32$ ,  $df = 1$ ,  $p = .250$ ) and fifth grade ( $\chi^2 = 3.48$ ,  $df = 1$ ,  $p = .062$ ).

## **Discussion**

The purpose of the present study was to compare the effectiveness of two variants of teacher training courses developed to gear reading comprehension instruction in elementary schools to the findings and research-based successful practices emanated

from empirical intervention research. More specifically, the efficacy of a year-round elaborated coaching of the teachers was equated to a more restricted 13-hours in-service course by using a questionnaire assessing teachers' experiences with the training course and with the implementation of the innovations, as well as student outcomes as a result of the renewed instructional practice.

### *Student Outcomes*

With regard to student outcomes the results of the multilevel linear regression analyses can be easily summarized: the two teacher training conditions were equally effective in changing students' reading comprehension, reading fluency, use of reading strategies, and self-efficacy perceptions towards reading from pre- to post- and retention test. These results confirm the hypothesis that student outcomes would be comparable in both teacher training courses.

### *Experiences Reported by the Teachers*

Overall, teachers in both the intensive year-round coaching and the restricted in-service training condition reported positive experiences. A picture emerges that second- and fifth-grade teachers were both strongly satisfied with their participation in the training course, with the offered information and counseling, and with the provided manual and student materials. Moreover, their reports indicate that implementing the innovative reading comprehension instruction as a result of attending one of the teacher training courses demanded only little to moderate additional workload. Teachers also perceived the organization of peer tutoring activities as a moderate to strong vehicle to build in differentiation according to pace or content for reading dyads of different abilities. As an additional advantage they report moderate to strong student improvement in the field of reading comprehension and fluency skills, social and interaction skills, social relationships and friendships, attitude towards reading, and responsibility, self-confidence, and autonomy when approaching assignments as a result of the implementation of the innovative reading comprehension instruction.

To investigate the differences in the experiences reported by the teachers in the different training conditions, analyses of variance were performed on the components of the teacher questionnaire. The results revealed that second- and fifth-grade teachers in both teacher training conditions were equally satisfied with the training course they attended and with the provided manual and student materials. Except for teachers' assessment of students' reading comprehension progress, no differences were detected either with regard to teachers' appraisal of students' progress.

As concerns the difference in the evaluation of students' reading comprehension progress, the results more specifically favored the restricted in-service teacher training condition, for teachers in this condition, especially those teaching fifth grade, reported larger student improvement than the intensively coached teachers. This result contrasts with the finding that with respect to the actual student learning gains no differences were found between the two strategies for professional development of teachers.

Presumably, the finding that teachers in the more restricted in-service training condition perceived higher learning progress in their students' reading comprehension can be attributed to the newness of the peer tutoring activities for those teachers, whereas teachers in the intensive coaching condition implemented the entire innovative approach (i.e., explicit reading strategies instruction and peer tutoring activities) for the second consecutive year, which could have led to a habituation effect. For teachers in the more restricted in-service training the implementation of peer tutoring activities in their classrooms was something completely new since the previous year they had only implemented part of the innovation, more particularly the explicit teaching of reading strategies. It is conceivable that the teachers' enthusiasm for this new experience led them to perceive greater student progress.

Finally, the analyses with regard to the perceived workload as a result of implementing the innovative instructional approach did reveal significant differences between both teacher training conditions. More specifically, it appeared that teachers attending the restricted in-service training course experienced a higher workload with regard to settling in the innovative approach and preparing the students for practicing the application of reading strategies in peer tutoring dyads. Apparently, this finding goes together with the fact that in the restricted in-service teacher training only three three-hours preparatory meetings with the teachers were organized, while in the intensive coaching condition eight local meetings took place. So, the former group of teachers needed more additional time apart from the in-service training sessions to study and prepare for the intended innovative instructional approach. It is very much conceivable that individual extra preparation time is perceived as causing more extra workload compared to the comfort of attending coaching sessions in which everything can be discussed with the external coach. But also the fact that for the teachers in the more restricted in-service training condition implementing peer tutoring was something new – while teachers in the intensive coaching condition had previous experience in it – could be responsible for this higher perception of workload.

Notwithstanding the higher report of pressure of preparatory work, fifth-grade teachers attending the restricted in-service training considered the organization of the peer tutoring activities more strongly as an effective opportunity to build in and increase differentiation according to pace or content for children of different abilities. On the basis of the content and structuring of both teacher training strategies, however, no obvious explanation of this difference can be formulated. But again, the fact that compared to the teachers in the other condition, implementing peer tutoring was something new for teachers in the restricted in-service training condition may have caused them to appreciate in greater extent some of the pedagogical advantages of this new teaching method.

In conclusion, the results concerning the experiences reported by the teachers generally confirm the research hypothesis that teachers' experiences in the restricted in-service course are comparable to those in the intensive coaching condition. For only three elements in the teacher questionnaire significant differences could be determined, of which two differences favored the restricted in-service course. The third difference suggests that teachers in the restricted in-service course need to spend more time and energy on settling in the innovative approach and preparing the

students. The additional workload, as perceived by teachers in the restricted in-service training does, however, not undermine teachers' satisfaction with attending the training course.

### *Conclusion*

The present study evaluated the effectiveness of two variants of teacher training courses, characterized by different amounts of support and assistance lent to the teachers, in making reading comprehension instruction in elementary schools more effective. More specifically, we intended to develop a concise teacher training program, of which the effects on teachers and student outcomes would be comparable with those of an intensive year-round counseling, coaching, and on-the-job-training of teachers. Therefore, the restricted in-service course was designed to include research-based effective components of teacher training.

The analyses on both the experiences reported by the teachers and student outcomes generally confirmed the central research hypothesis and clearly corroborated that a compact in-service teacher training aimed at changing instructional content and teacher behaviors in the classrooms was equally effective as the year-round coaching of teachers. Taken into account previous studies referred to above, the similar effectiveness of the restricted in-service teacher training as compared to the intensive year-round coaching can be attributed to the combination of different components, notably the focused presentation of information about the background of the innovative instructional approach, the identification and discussion of the intended instructional practice, the feasible and elaborated manual with lesson scenarios and all necessary teacher and student materials, the demonstration of relevant instructional behaviors to be implemented by means of accurately selected video fragments, and follow-up sessions with in-class supervised practice, on-site consultation, and performance feedback to the teachers.

Finally, some limitations of the present study should be noted. First, it should be mentioned that teachers were asked to deliver or mail the questionnaire with regard to their experiences with the training program to the training course instructor. Therefore, it is possible that socially desirable responses were elicited from some of the respondents.

A second restriction is connected to the fact that the author of the article who conducted both teacher training programs was already experienced in intensively coaching teachers to innovate their reading comprehension instruction from former research (Van Keer, 2002; Van Keer & Verhaeghe, 2002a, b). In this respect, the objection can be raised that because of the previous elaborated coaching and counseling experience, the quality of the restricted in-service teacher training was higher than can be expected in the case of a novice or less experienced counselor of the teacher training sessions. Further research is necessary to investigate whether the quality of the present 13-hours in-service teacher program and the connected teacher satisfaction can be equaled by disseminating the innovative approach on a broad scale through regular in-service training authorities. Since these authorities often prefer to

assemble training participants from different school teams, additional research should in this respect also try to verify whether the organization of the three three-hours preparatory local school meetings can be replaced by collective meetings without losing quality.

A third comment can be made on the fact that only experiences reported by the teachers and student outcome measures were used to compare the effectiveness of both teacher training programs. It can be argued that it is as important to gather information about the extent to which the implementation of the innovations has been carried out correctly and in accordance with the intention of the teacher training program. Quality of the implementation data were, however, not collected systematically. Therefore, subsequent research should complete the data gathered in the present study with evaluations of the treatment integrity of the innovative reading comprehension instruction approach as a result of the different teacher training programs.

## References

- Aarnoutse, C. A. J. (1995). *Probleemidentificatie en aanzet voor een actieplan taal* [Problem identification and initiative towards an action plan for language]. Nijmegen, The Netherlands: Vakgroep Onderwijskunde.
- Aarnoutse, C. A. J., & Weterings, A. C. E. M. (1995). Onderwijs in begrijpend lezen [Education in reading comprehension]. *Pedagogische Studiën*, 72, 82-101.
- Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.
- Alvermann, D. E. (2000). Classroom talk about texts: Is it dear, cheap or a bargain at any price? In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 170-192). New York: Teachers College Press.
- Ball, D. L., & Cohen, D. (1996). Reform by the book: What is - or might be - the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25, 6-8.
- Bentz, J. L., & Fuchs, L. S. (1996). Improving peers' helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly*, 19, 202-215.
- Brand-Gruwel, S., Aarnoutse, C. A. J., & Van Den Bos, K. P. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and instruction*, 8, 63-81.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.
- Brus, B. Th. (1969). *Eén-minuut-test* [One minute test]. Nijmegen, The Netherlands: Berkhout.
- De Corte, E., Verschaffel, L., & Van De Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Educational Psychology*, 71, 531-559.

- Djalil, A., & Anderson, L. W. (1989). The impact of a research-based teacher training program on Indonesian teachers, classrooms and students. *Teaching and Teacher Education*, 5, 165-178.
- Dole, J. A. (2000). Explicit and implicit instruction in comprehension. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 52-69). New York: Teachers College Press.
- Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61, 239-264.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliff, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putnam, J., & Bassari, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22, 347-368.
- Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, C. L. (1994). The nature of students interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75-103.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K. & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 97, 223-249.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [Step out! Manual]. Utrecht, The Netherlands: Sardes.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Haller, E. P., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher*, 17, 5-8.
- Harchik, A. E., Sherman, J. A., Sheldon, J.B., & Strouse, M.C. (1992). Ongoing consultation as a method of improving performance of staff members in a group home. *Journal of Applied Behavior Analysis*, 25, 599-610.
- Hawley, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as a learning profession: Handbook of policy and practice* (pp. 127-150). San Francisco: Jossey-Bass.
- Joyce, B., & Showers, B. (1980). Improving in-service training: The message of research. *Educational Leadership*, 37, 379-385.
- Joyce, B., & Showers, B. (1982). The coaching of teaching. *Educational Leadership*, 40, 4-10.
- Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, 96, 275-293.



- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, 99, 3-21.
- Leach, D. J., & Ingram, K. L. (1989). The effects of information and feedback on teachers' classroom behavior and students' academic engaged time. *Educational Psychology*, 9, 167-184.
- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.
- McCutchen, D., & Berninger, V. W. (1999). Those who know, teach well: Helping teachers master literacy-related subject-matter knowledge. *Learning Disabilities Research and Practice*, 14, 215-226.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92, 513-555.
- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal*, 90, 3-32.
- Pressley, M., Wharton-McDonald, R., Hampston, J. M., & Echevarria, M. (1998). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies of Reading*, 2, 159-191.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Showers, B. (1990). Aiming for superior classroom instruction for all children: A comprehensive staff development model. *Remedial and Special Education*, 11, 35-39.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Staphorsius, G., & Krom, R. (1996). *Toetsen Begrijpend Lezen* [Reading Comprehension Tests]. Arnhem, The Netherlands: Cito.
- Van Keer, H. (in press). *Een boek voor twee. Verwerving van strategieën voor begrijpend lezen via peer tutoring*. [A book for two. Acquisition of reading comprehension strategies through peer tutoring.] Leuven, Belgium: Garant.
- Van Keer, H. (2002). *Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring*. Manuscript submitted for publication.
- Van Keer, H., & Verhaeghe, J. P. (2002a). *Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions*. Manuscript submitted for publication.

- Van Keer, H., & Verhaeghe, J. P. (2002b). *Effects of peer tutoring on second graders' reading fluency development*. Manuscript submitted for publication.
- Veenman, S., Van Tulder, M., & Voeten, M. (1994). The impact of in-service training on teacher behavior. *Teaching and Teacher Education, 10*, 303-317.
- Verhoeven, L. (1993). *Lezen met begrip 1* [Reading with comprehension 1]. Arnhem, The Netherlands: Cito.
- Wade, R. K. (1985). What makes a difference in in-service teacher education? A meta-analysis of research. *Educational Leadership, 42*, 48-54.
- Wheldall, K., Merrett, F., & Borg, M. (1985). The behavioural approach to teaching package (BATPACK): An experimental evaluation. *British Journal of Educational Psychology, 55*, 65-75.

## Chapter 6

### Effects on second and fifth graders' reading comprehension, strategy use and self-efficacy related thoughts: Results of the second study\*

#### Abstract

This study investigated the effects of explicit reading comprehension strategies instruction, followed by practice in either reciprocal same-age (STRAT+SA) or cross-age peer tutoring dyads (STRAT+CA) on second and fifth graders' reading comprehension, strategy use, and self-efficacy perceptions. For second grade multilevel analysis following a repeated measures design revealed significant positive posttest effects for reading comprehension (STRAT+CA), strategy use (both experimental conditions) and reports of positive thoughts related to reading proficiency (both experimental conditions). Significant positive retention test effects were found for strategy use (STRAT+CA). In fifth grade significant positive posttest effects on reading comprehension were found for both experimental conditions and a significant positive retention test effect for the STRAT+CA condition only. A significant decrease in negative thoughts related to reading proficiency was found in STRAT+SA fifth graders at the posttest.

#### Introduction

The present article describes the results of a study designed to evaluate a complex set of instructional interventions for the acquisition of strategic reading comprehension skills as applied to regular elementary classrooms. To develop the interventions, we drew on two research fields, namely explicit reading comprehension strategy instruction (e.g., Brown, Pressley, Van Meter, & Schuder, 1996; De Corte, Verschaffel, & Van De Ven, 2001; Klingner, Vaughn, & Schumm, 1998; Palincsar & Brown, 1984, Pressley et al. 1992) and peer tutoring (e.g., Bentz & Fuchs, 1996; Cohen, Kulik, & Kulik, 1982; Fantuzzo, Polite, & Grayson, 1990; Fuchs, Fuchs, Mathes, & Simmons, 1997; Mathes & Fuchs, 1994; Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995; Topping, 1996). More specifically, the purpose of the research was to explore the efficacy of the innovative instructional approach as a tool for fostering second and fifth graders' use of reading strategies and reading comprehension achievement and for ameliorating their self-efficacy perceptions towards reading.

Further in this introductory section the concepts of reading comprehension and strategic reading will be presented. Thereafter, an overview of research evidence, revealing the importance of explicit reading strategies instruction and interaction among peers to enhance strategic reading and comprehension, will be summarized. Finally, the concept of peer tutoring, as a potentially effective instructional technique to stimulate peer-led interaction about texts, will be described.

---

\* Based on: Van Keer, H., & Verhaeghe, J. P. Strategic Reading in Peer Tutoring Dyads in Second- and Fifth-Grade Classrooms. The present chapter is submitted for publication in *The Elementary School Journal*.

Reading comprehension can be defined as constructing a mental representation of textual information and its interpretation (Van Den Broek & Kremer, 2000) and refers to understanding and putting a meaning on written words, sentences, and texts (Aarnoutse & Van Leeuwe, 1998, 2000). Previously, comprehension was assumed to occur automatically once students could decode (Dole, 2000). From this perspective, the simple view of reading claims that reading comprehension merely depends on two components, namely word recognition and linguistic comprehension (i.e., vocabulary and listening comprehension) (Hoover & Gough, 1990). Research, however, shows that reading comprehension is more than being able to decode and listening to what has been read to oneself. Reading comprehension is an active process and is affected by continuous complex interactions between the reader and the text (Paris, Wasik, & Turner, 1991) and by various cognitive and metacognitive processes (Pressley & Afflerbach, 1995). Cognitively based views of reading comprehension emphasize that proficient readers do much more than just word-, phrase-, or sentence-level processing (Dole, Duffy, Roehler, & Pearson, 1991); the mastery and use of both metacognitive and cognitive strategies that facilitate text comprehension distinguishes poor and skilled readers (Brand-Gruwel, Aarnoutse, & Van Den Bos, 1995; Palincsar & Brown, 1984).

Cognitive reading strategies can be defined as a range of mental and behavioral activities, such as rereading, activating prior background knowledge, and adjusting reading speed, used to increase the likelihood of comprehending text and to handle comprehension failures (Van Den Broek & Kremer, 2000). These strategies are conscious, instantiated, and flexible plans readers apply and adapt deliberately to a variety of texts and tasks (Baker & Brown, 1984; Paris et al., 1991; Pressley & Allington, 1999; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989). On the other hand, metacognitive strategies can be specified as ongoing self-monitoring and self-regulating activities supporting readers' awareness of (loss of) comprehension and assisting their decision of what cognitive strategies to employ to aid comprehension as a function of text difficulty, situational constraints, and the reader's own cognitive abilities (Lories, Dardenne, & Yzerbyt, 1998; Van Den Broek & Kremer, 2000; Weisberg, 1988).

The mastery of cognitive and metacognitive strategies is critical to become a skilled reader. Nevertheless, not all elementary students spontaneously develop both the metacognitive skills to monitor and regulate comprehension and the "fix up" cognitive strategies to repair understanding when it breaks down (Hartman, 2001; Pressley & Allington, 1999). Evidence suggests, however, that explicit teacher-mediated strategies instruction can effectively promote strategic reading and reading comprehension (e.g., Brand-Gruwel et al., 1998; De Corte et al., 2001; Dole et al., 1991; Duffy et al., 1987; Haller, Child, & Walberg, 1988; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984; Pressley, 2000; Pressley et al., 1989; Walraven, 1995). Unfortunately, this type of reading comprehension instruction is not offered in many classrooms. The instructional practice of teaching reading in elementary schools is mainly characterized by questioning students about the content of a text after reading it, with hardly any overt and continuous instruction in strategy selection, use, and evaluation (Aarnoutse, 1995; Aarnoutse & Weterings, 1995; Dole, 2000; Durkin,

1978-1979; Paris & Oka, 1986; Pressley, Wharton-McDonald, Hampston, & Echevarria, 1998; Weterings & Aarnoutse, 1986).

Besides the significance of explicit reading strategies instruction, the effectiveness of opportunities to negotiate meaning and to participate in peer-led interaction on structured reading activities has also been documented (e.g., Almasi, 1996; Brown et al., 1996; Dole et al., 1991; Fuchs, Fuchs, Mathes, et al., 1997; Greenwood, Delquadri, & Hall, 1989; Johnson-Glenberg, 2000; Mathes, & Fuchs, 1994; Mathes, Torgesen, & Allor, 2001; Palincsar & Brown, 1984; Pressley et al., 1992; Rosenshine & Meister, 1994; Simmons et al., 1995). More particularly, evidence suggests that through discussions, peer conferences, peer tutoring, and cooperative activities students implement, evaluate and modify strategy acquisition and use, and discuss how a strategy could be applied in situations other than the reading lessons (Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984). Moreover, discussions between peers provide opportunities for metacognitive exchanges and modeling (Palincsar, David, Winn, & Stevens, 1991). Notwithstanding these convincing research results, student-centered discussion with regard to reading comprehension is anything but common practice in most classrooms (Alvermann, 2000).

Taken into account the yawning gap between the educational practice of reading comprehension instruction on the one hand and the research evidence demonstrating the importance of explicit attention to the strategic aspects of text processing and comprehension and peer-led interaction about texts on the other hand, an intervention study was set up to evaluate an innovative approach aimed at making reading comprehension instruction in elementary schools more effective. More specifically, the interventions combine explicit reading strategies instruction with peer tutoring, as an instructional technique to create opportunities to stimulate student-led interactions about texts.

Peer tutoring can be defined as "people from similar social groupings who are not professional teachers helping each other to learn, and learning themselves by teaching" (Topping, 1996, p. 322). In peer tutoring, two important categories can be distinguished with regard to the composition of the dyads. In same-age tutoring children are paired with other children from within their own classroom. The specific form of same-age tutoring in which the students alternate on a regular basis between the tutor and tutee role is called reciprocal same-age tutoring (Fantuzzo, King, & Heller, 1992). The second variant is called cross-age tutoring and refers to older students tutoring younger students. In this case role reciprocity is generally not applicable. In comparison with teacher-directed instruction, peer tutoring has been found to increase individualization, time on task, immediate and specific feedback, reinforcement, and error correction, as well as opportunities to respond, academic engagement, and relevant academic behaviors (e.g., Greenwood, Carta, & Hall, 1988; Greenwood & Delquadri, 1995; King-Sears & Bradley, 1995; Simmons et al., 1995; Utley & Mortweet, 1997).

Researchers have documented the effectiveness of peer tutoring for a variety of curriculum areas and different age groups. Positive research results were found on academic achievement for both tutor and tutee (e.g., Cohen et al., 1982; Fantuzzo et al., 1990; Fantuzzo et al., 1992; Greenwood, Terry, Arreaga-Mayer, & Finney, 1992; Greenwood, Terry, Utley, Montagna, & Walker, 1993), in particular for children with learning problems (e.g., Bentz & Fuchs, 1996; Fuchs, Fuchs, Mathes, et al., 1997; Klingner & Vaughn, 1996; Mathes & Fuchs, 1994; Simmons et al., 1995). Positive research results were also found on tutors' and tutees' social, and emotional functioning, more specifically with regard to improved self-efficacy perceptions, more positive self-concepts and social relationships, and better attitudes towards the curriculum areas treated in the tutoring sessions (e.g., Cohen et al., 1982; Fantuzzo et al., 1992; Fantuzzo, Davis, & Ginsburg, 1995; Greenwood et al., 1988; Mathes & Fuchs, 1994; Roswal et al., 1995). As concerns the differential impact of same-age and cross-age tutoring, a previous comparative study (Van Keer & Verhaeghe, 2002) reveals larger effect sizes for both second- and fifth graders in a cross-age tutoring program. However, no statistically significant differences between the variants were found.

### **Study's Purpose and Hypotheses**

The aim of the present study was to implement and examine the effectiveness of an innovative approach to reading comprehension instruction, which combines explicit reading strategies instruction with strategic reading practice in either cross-age (STRAT+CA) or reciprocal same-age peer tutoring activities (STRAT+SA), by analyzing the results of reading comprehension, strategy use, and self-efficacy perceptions towards reading as outcome measures.

The present study is part of a larger long-term study investigating the effects of a program innovating reading comprehension instruction in Flanders and can be considered as a partial replication of a previous research during the school year 1999-2000 (Van Keer, 2002; Van Keer & Verhaeghe, 2002). Due to the scope and the means of the study, the data of the first study's control group were also used in the present study. The implementation of the experimental interventions occurred with a second cohort second and fifth graders and was spread out over the entire school year 2000-2001 with the participation of all students.

In comparison with the previous evaluation of the STRAT+SA and STRAT+CA interventions (Van Keer, 2002; Van Keer & Verhaeghe, 2002) twice as much classes were included in both the STRAT+SA and STRAT+CA conditions with a view to verify the research results found before with a larger sample. Moreover, the present study extends other studies by (a) enabling an explicit comparison of the potentially differential impact of same-age and cross-age peer tutoring activities; (b) supporting teachers to implement the innovations in their regular classroom context with the participation of all students in the course of an entire school year; (c) including long-term maintenance assessments; (d) using standardized reading comprehension tests not linked directly to the treatment; and (e) applying multilevel modeling to take the hierarchical nesting of students in classes into account. Based on a review of the

relevant research literature, the major hypotheses of the study can be formulated as follows:

1. Both the STRAT+SA and STRAT+CA experimental interventions enhance second and fifth graders' reading comprehension achievement, strategy use, and self-efficacy perceptions towards reading more than traditional reading comprehension instruction. Therefore, we expect a significantly higher progress from pretest to post- and retention test on the three dependent measures for the students in the STRAT+SA and STRAT+CA condition as compared to the students in the control group.
2. The improvement in reading comprehension, strategy use, and self-efficacy perceptions towards reading is more obvious for second graders functioning as tutees in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities. This prediction is tested by pairwise comparisons of the growth of second-grade STRAT+SA students with that of STRAT+CA students.
3. The improvement in reading comprehension, strategy use, and self-efficacy perceptions towards reading is more obvious for fifth graders functioning as tutors in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities. This hypothesis is tested by pairwise comparisons of the STRAT+SA fifth graders' progress on the one hand and the and STRAT+CA students' progress on the other hand.

### **Description of the Innovative Approach**

This section shortly outlines the innovative approach implemented within the framework of the present study. A more thoroughly description can be found in previous articles (Van Keer, 2002; Van Keer & Verhaeghe, 2002) or in the manual (Van Keer, in press). The approach is characterized by three elements, namely explicit reading strategies instruction, a sound tutor preparation, and practice of reading strategies in weekly cross-age or reciprocal same-age peer tutoring sessions. Two distinguished versions of the innovative approach were tested. One combines explicit reading strategies instruction with reciprocal same-age peer tutoring; the other combines explicit reading strategies instruction with cross-age tutoring. All classroom activities that were part of the experimental treatment were conducted by the regular classroom teachers who got substantial support and training. In this respect it should be noted that the interventions were not meant as an additional program on top of teachers' traditional reading comprehension classes. Experimental teachers were coached to substitute their traditional way of teaching reading comprehension and they were encouraged to implement the treatments during time normally allocated for reading instruction.

### *Explicit Instruction in Reading Strategies*

The major changes due to the explicit reading strategies instruction related to the content of teaching and learning, and to the instructional techniques. In terms of the content, the innovation focused on the acquisition and mastery of six reading strategies, namely: (a) activating prior background knowledge and linking it to the text, (b) predictive reading and verifying story outcomes, (c) distinguishing main issues from side-issues, (d) monitoring and regulating the understanding of words and expressions, (e) monitoring and regulating comprehension by tracing the ideas expressed in difficult and not understood sentences or passages, and (f) classifying types of text and adjusting reading behavior to it. The selection of the strategies was mainly based on an extensive review of related empirical research and innovation programs focusing on explicit strategy instruction (e.g., Brown et al., 1996; De Corte et al., 2001; Fuchs, Fuchs, Mathes, et al., 1997; Fukkink, Van der Linden, Vosse, & Vaessen, 1997; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984).

As concerns the instructional techniques, the innovative approach was influenced primarily by studies on transactional strategies instruction (Brown et al., 1996; Pressley et al., 1992) and reciprocal teaching (Klingner & Vaughn, 1996; Palincsar & Brown, 1984). In this respect, much attention was paid to extensive and explicit teacher explanations and modeling of strategic reasoning. The explicit instruction involves the intended, direct, and systematic presentation of information by the teacher to the students (Simmons et al., 1995). Within the framework of the study, this implied that children were instructed in strategies which support comprehension; in why, where, and when to use them; as well as in how to adapt them to various situations. In addition, a gradual transfer from external regulation by the teacher to self-regulation of strategy use by the students was taken into account, for a long-term goal is the self-regulated use of strategies.

### *Tutor Preparation*

Since there is evidence that peer tutoring is less effective when no attention is paid to a sound training of the tutors (e.g., Bentz & Fuchs, 1996; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997), a series of preparatory lessons and materials was developed, based on related research and tutoring programs (Bentz & Fuchs, 1996; Fukkink et al., 1997). Verbal and visual explanations, modeling, role plays, discussions, and student practice with teacher feedback represented an important part of the training. The preparatory lessons were scheduled at the beginning of the intervention and required seven 50-minute sessions. Tutors more particularly got acquainted with their tasks and responsibilities, they learned how to show interest, how to initiate and finish a session, how to give corrective feedback, and offer positive reinforcement for correct answers. Moreover, students were introduced to guidelines for knowing when and how to offer appropriate explanations, help, and assistance.



### *Peer Tutoring Sessions*

Finally, the innovative approach was typified by weekly peer tutoring activities to systematize the use of the explicitly taught reading strategies. Peer tutoring activities were organized class-wide, so all students in a class were paired and worked simultaneously. Some second- and fifth-grade teachers implemented cross-age tutoring activities, others implemented reciprocal same-age tutoring activities. In the cross-age activities fifth graders were paired with second-grade students. In the reciprocal same-age peer tutoring activities second and fifth graders were paired with classmates. Students in the same-age dyads alternated regularly between tutee and tutor roles, allowing each individual to take turns acting as the dialogue leader. Teachers saw to it that both students served as tutor for an equal amount of time.

Teachers were responsible for organizing the tutoring sessions; for observing, monitoring and supporting the dyads; and for providing corrective and instructional feedback and praise. Peer tutoring sessions were organized once or twice a week on the average and lasted from 25 to 50 minutes, depending on the task and scheduling. With regard to the succession of the introductory strategy instruction and the more independent practice in peer tutoring sessions, a "sandwich model" was applied. This implies that the teacher-led introduction of a new reading strategy was followed by at least one peer tutoring session in which the strategy was practiced when reading one of the selected text excerpts from the teachers' manual. Thereafter a number of tutoring sessions followed in which the strategy was practiced while reading books or texts the students in the reading dyads could choose themselves from the school or class library. Taken into account prior research results stressing that a combination of dyadic peer interaction and structured academic activity do more to enhance cognitive gain than either of the two dimensions separately (Cohen et al., 1982; Fantuzzo, Riggio, Connelly, & Dimeff, 1989; Lambiotte et al., 1987), strategy assignments cards were used as a vehicle for structuring the peer tutoring interaction about the application of the reading strategies.

### *Support to the Teachers*

Because the interventions were implemented by the regular classroom teachers, teachers were prepared and supported thoroughly to ensure that the innovations were administered with fidelity. All participating teachers were provided with an elaborated manual (Van Keer, in press) including all materials necessary to conduct the innovation, namely (a) an extensive general description of the rationale, the aims, and the organization of the interventions, (b) lesson scenarios structuring the implementation and describing the objectives, the necessary materials, the preferable instructional techniques, and the successive phases of each lesson, and (c) supplementary students materials, such as strategy assignment cards and reading texts. The manual was based on extensive review of related empirical research (e.g., Bentz & Fuchs, 1996; Brown et al., 1996; Fuchs, Fuchs, Mathes, et al., 1997; Fukkink et al., 1997; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984; Pressley et al., 1992) and prevailing manuals focusing on explicit reading comprehension instruction (e.g., Aarnoutse & van de Wouw, 1990).

In addition to the manual, teachers were provided with in-service training and technical assistance.

## Method

### *Design*

The effectiveness of the innovations was tested in a quasi-experimental study with a pretest posttest and retention test control group design. More specifically, classes were assigned to three conditions: (a) explicit reading strategies instruction and practicing strategic reading in reciprocal same-age peer tutoring activities (STRAT+SA), (b) explicit reading strategies instruction and practice in cross-age peer tutoring activities (STRAT+CA), and (c) traditional reading comprehension instruction without explicit instruction and peer tutoring or other forms of peer-led interaction (control group). Control group teachers were told that the purpose of the study was to examine elementary school students' development in the field of reading comprehension and were not informed that they were part of a control condition. Interviews with these teachers revealed that the reading comprehension lessons typically involved teacher-led whole-class activities, including asking comprehension-check questions after reading a text, teacher evaluation of students' answers, and presentation of the correct answers.

Given that intact, naturally assembled classes were assigned to the research conditions, the design can be referred to as quasi-experimental. Because the difficulty often encountered when attempting to form groups by random assignment, quasi-experimental research is quite common in education (Wiersma, 2000). This setting ensures the ecological validity of the interventions and provides a more stringent test of the successful implementation of the interventions than studies in tightly controlled laboratory settings, of which research results can not simply be transferred to the context of real-life classrooms.

### *Participants*

Participants included all of the students in 20 second- and 22 fifth-grade classrooms from 19 different schools throughout Flanders (Belgium). In total 396 second graders and 449 fifth graders participated. The schools' population was comprised chiefly of white, Flemish students from middle-class families, with the exception of one inner-city school (one STRAT+CA second- and fifth-grade class) with mainly a low SES and ethnic minority population. Except for one fifth-grade class including only boys, there was approximately an even division of gender: on average 53% of the students were boys in both the second- ( $SD = 15.85$ ) and fifth-grade classes ( $SD = 17.61$ ). The majority of the students (352 in second and 419 in fifth grade) were native Dutch speakers, which is the medium of instruction in Flanders. Classes are to be considered as academically heterogeneous. Class size ranged from 10 to 26 with an average of 20

( $SD = 3.13$ ) in second grade and from 9 to 34 in fifth grade with on average 22 ( $SD = 6.40$ ) students per class, which is representative for the Flemish situation.

Participating teachers were selected two years before from a group of about 100 second- and fifth-grade teachers, all willing to take part in a long-term research study. The selection was based on the geographical distribution of the schools throughout the whole of Flanders on the one hand, and on the possibility to match teachers and classes as closely as possible with regard to teachers' teaching experience, class size, students' age, gender distribution, and dominating mother tongue on the other hand. Within the peer tutoring conditions teachers were allowed to opt in favor of the STRAT+SA or STRAT+CA condition according to their preference and the readiness of a second- or fifth-grade colleague to collaborate in the cross-age activities. The experimental teachers in the present study participated for the second year. In the year before approximately half of the teachers had participated in either the specific STRAT+SA or STRAT+CA condition. The other half had been involved in an experimental condition characterized by explicit reading strategies instruction and practicing strategic reading in teacher-led whole-class activities. Control group classes were selected to match the teachers and class groups in the experimental conditions. Table 6.1 shows the number of classes and students in each condition for the present study.

Table 6.1  
*Number of Participating Classes and Students*

Condition	Grade			
	2 <sup>nd</sup>		5 <sup>th</sup>	
	Classes	Students	Classes	Students
STRAT+SA	6	110	9	186
STRAT+CA	8	162	7	156
Control group	6	124	6	107
Total	20	396	22	449

### *Measurement Instruments*

Standardized reading comprehension tests, as well as questionnaires assessing the use of reading strategies and preoccupation with attributions and self-efficacy perceptions towards reading were administered as dependent measures.

#### *Standardized Reading Comprehension Tests*

In second grade reading comprehension achievement was measured using well-established and widely used Dutch standardized test batteries. At each measurement occasion a different test, increasing in level of difficulty, was used. For the pretest a test from the battery "Lezen met begrip 1" (Reading with comprehension) (Verhoeven, 1993) was utilized. For the post- and retention test, tests from the battery "Toetsen Begrijpend Lezen" (Reading comprehension tests) (Staphorsius & Krom, 1996) were administered. After discussing an example, students completed the tests individually. The students read short expository and narrative texts and then answered multiple choice questions with four alternatives. The questions pertain to the word, sentence, or

text level. The administration of each test lasted about one hour. The scores are determined by summing up the correct answers. The Cronbach's  $\alpha$ -coefficients yielded high reliability scores of .90 ( $n = 390$ ) for the pretest (30 items), .82 ( $n = 382$ ) for the posttest (25 items), and .82 ( $n = 348$ ) for the retention test (25 items).

In fifth grade reading comprehension achievement was measured with the standardized test battery "Toetsen Begrijpend Lezen" (Reading comprehension tests) (Staphorsius & Krom, 1996). Different tests, increasing in level of difficulty, were used for the three measurement occasions. Two types of questions, both requiring the integration of information on different textual levels, can be distinguished, notably questions concerning the content of a text and concerning the communication between the author and the reader. After discussing an example, students completed the tests individually. The tests consist of three modules of 25 multiple-choice questions with four alternatives each. All students took the first module of the test. Afterwards they completed the second more easy or third more difficult module, depending on their first results. The scores are determined by summing up the correct answers. To compare the scores of the students who finished the more easy or more difficult part of the test, the sum scores were transposed into an IRT-modeled global achievement score following the test guidelines. This score ranges between 0 and 100, with the values 0, 50 and 75 showing a reading comprehension achievement respectively far below the average achievement in second grade, around the average at the end of fifth grade and above the average achievement in sixth grade. To verify the reliability of the three modules of the pre-, post-, and retention test, Chronbach's  $\alpha$ -coefficients were computed. Table 6.2 indicates that all reading comprehension measures are reliable.

Table 6.2

*Chronbach's  $\alpha$ -coefficients of the Fifth-Grade Reading Comprehension Tests*

Test module	Measurement occasion <sup>a</sup>		
	Pretest	Posttest	Retention test
1	.80 ( $n = 443$ )	.73 ( $n = 422$ )	.76 ( $n = 378$ )
2	.75 ( $n = 154$ )	.82 ( $n = 243$ )	.78 ( $n = 346$ )
3	.70 ( $n = 276$ )	.74 ( $n = 166$ )	.70 ( $n = 31$ )

<sup>a</sup>At each measurement occasion a different test, increasing in level of difficulty, was used.

### *Questionnaire on the Use of Reading Comprehension Strategies*

To estimate to what extent second- and fifth graders use reading comprehension strategies while reading, a questionnaire was developed as part of the present study. More specifically, it concerns a list of 20 statements with respect to the use of strategies before, while, and after reading. The students were asked to report how often the statements applied to their own reading behavior by ticking one of the boxes ("never", "almost never", "sometimes", "very often"). Examples of the statements are: "Before I start reading, I consider what I already know about the subject of the text.", "While reading, I look for the meaning of difficult words", "While reading, I try to figure out what the story is about." The questionnaire was administered at each measurement occasion to both second and fifth graders. Fifth graders read and

completed the questionnaire individually. In second grade all items were read out loud and judged individually by the students. Chronbach's  $\alpha$ -coefficients are presented in Table 6.3 and reveal that the questionnaire is on an acceptable level of reliability for both second and fifth graders.

### *Questionnaire on Self-efficacy Perceptions and Related Causal Attributions*

A questionnaire was developed to measure how often students are preoccupied with positive or negative thoughts with regard to their own reading ability or related causal attributions. Inspired by the work of Ames (1984), children were asked to report how often such thoughts crossed their mind either before, while, or after reading by ticking off "never", "almost never", "sometimes", or "very often", following statements as: "I think: I comprehended well because it was an easy text.", "I think: I did not comprehend well because there was no one to help me.", "I think: I am not good at reading". Factor analysis revealed that success attributions and positive thoughts about one's own reading competence on the one hand and failure attributions and negative self-efficacy perceptions with regard to reading on the other hand are very closely related. This result is in line with the findings of Marsh (1984) and Marsh, Cairns, Relich, Barnes, and Debus (1984) that self-attributions can be seen as expressions or indicators of one's self-concept or self-efficacy perceptions. Therefore, two scales were constructed. It is important to mention that a low score on the scale with regard to positive thoughts relating to self-efficacy does not necessarily mean that a student has a low self-esteem with regard to reading proficiency. It only reveals that the student is not preoccupied with thoughts about his reading proficiency or success. On the other hand, a high score on the scale with regard to negative thoughts relating to self-efficacy in reading clearly indicates the presence of a low self-esteem with regard to reading. The questionnaire was administered at each measurement occasion to both second and fifth graders. Fifth graders read and completed the questionnaire individually. In second grade all items were read out loud by the first author and judged individually by the students.

Table 6.3 presents the results of the internal consistency analyses, revealing that both scales are on an acceptable level of reliability.

Table 6.3

*Chronbach's  $\alpha$ -coefficients of the Questionnaire concerning Preoccupation with Attributions and Self-efficacy Perceptions*

Questionnaire scale	Measurement occasion					
	Pretest		Posttest		Retention test	
	2 <sup>nd</sup> grade	5 <sup>th</sup> grade	2 <sup>nd</sup> grade	5 <sup>th</sup> grade	2 <sup>nd</sup> grade	5 <sup>th</sup> grade
Use of Reading Comprehension Strategies	.70 (n = 328)	.81 (n = 374)	.80 (n = 321)	.87 (n = 362)	.76 (n = 275)	.87 (n = 337)
Success attributions and positive self-efficacy perceptions	.62 (n = 364)	.66 (n = 421)	.72 (n = 371)	.68 (n = 402)	.66 (n = 322)	.67 (n = 360)
Failure attributions and negative self-efficacy perceptions	.76 (n = 345)	.82 (n = 387)	.79 (n = 343)	.86 (n = 375)	.83 (n = 297)	.85 (n = 335)

### *Decoding Fluency Test*

The "Eén minuut test" (One minute test) (Brus, 1969) was administered individually to all second graders at each measurement occasion to assess their reading fluency, which is a combination of accuracy and decoding speed (Chard, Simmons, & Kameenui, 1998). In this standardized test students read unrelated words with an increasing level of difficulty out of a list of 100 unrelated words during exactly one minute. The score is determined by counting the number of words read correctly.

### *Data Collection*

Data were collected within the regular classroom context and during regularly scheduled class sessions. All conditions were measured at three points in time, that is a pretest in October (second and fifth grade) before the initiation of the intervention program, a posttest in the period of May and June (second and fifth grade) after the completion of the experimental intervention, and a retention test in December (third and sixth grade). It is to be stressed that, in accordance with the planning of the research, none of the third- and sixth-grade teachers pursued the experimental intervention. Consequently, between post- and retention test (i.e., in the first term of third and sixth grade) all participants got traditional reading instruction and no experimental intervention was organized.

### *Data Analysis*

The present study has a clear hierarchical structure with students nested within a smaller number of classes. Moreover, the aim of the present study was to affect second- and fifth graders' reading comprehension performance, strategy use, and self-efficacy perceptions toward reading by interventions assigned at the level of the classroom. This implies that the experimental variable, assignment to one of the research conditions, is a group level variable, whereas the dependent variables (i.e., reading comprehension, strategy use, and self-efficacy perceptions toward reading) are measured at the individual level (Krull, 1999). Because of this joint modeling of individual and group variables, we take a multilevel modeling perspective on analyzing the data, for these models are specifically geared to the statistical analysis of data with a clustered structure (Goldstein, 1995). As Plewis and Hurry (1998) assert, the analysis of the data generated within the design adopted in the present study needs to be located within a multilevel framework to test the effectiveness of the educational interventions. They note:

Not only is this important in order to get an adequate representation of the precision of any intervention effect, it also offers various possibilities for investigating how the intervention effects operates, which go beyond a simple comparison of means across the groups. (pp. 24-25)

Within the framework of the present study, a repeated measures design was adopted, in which the measurement occasions were considered as a distinct level within the students. Consequently, a three-level hierarchical structure arises: the three measurement occasions (level 1) are clustered within students (level 2), in their turn nested within classes (level 3). Since data collection did not take place at identical time intervals, the measurement occasions were included in the model as dummies.

For each dependent variable, a stepwise procedure was followed to build appropriate models and to test the research hypotheses. The first step concerned the estimation of an unconditional three-level null model, without explanatory variables. This model served as a baseline with which to compare subsequent more complex models and partitioned the total variance of the dependent variables into three components: between-classes, between-students, and between-measurements variance.

Next, to gain a clear insight into students' general change in reading comprehension, strategy use, and self-efficacy perceptions from pretest to respectively post- and retention test, a compound symmetry model, which is a random intercept model with no explanatory variables except for the measurement occasions (Snijders & Bosker, 1999), was estimated. Therefore two dummies (i.e., post- and retention test contrasted against the pretest measure) were created and added to the fixed part of the model. To verify whether classes, individuals within these classes or both experience different changes in the course of time, the estimates of both dummies were allowed to vary randomly across all classes and students, yielding a fully multivariate model with regard to the repeated measures (Snijders & Bosker, 1999). Note that having decided that at the student level the effects of the measurement occasions could be random, it is not possible to estimate a random effect for the constant term at level one, for this coefficient is redundant in a fully multivariate model and is estimated as zero.

The third step consisted in the input of conceivably relevant explanatory variables, such as gender, mother tongue, and the number of years behind at school. Initially, these were included as fixed effects, assuming that the relation between the response and explanatory variables is constant across students and classes. Next, this assumption was relaxed by allowing the coefficients to vary randomly. The class-level total amount of time spent on reading comprehension instruction was also introduced as a potential predictor. However, there was evidence from these analyses that none of the response variables was significantly influenced by this variable.

Finally, with a view to test the postulated hypotheses, the effects attributable to the experimental interventions were explored. Therefore, the fourth step of the analysis consisted in adding the categorical variable "condition" to the model. To represent the research conditions two dummy-variables were used, contrasting the STRAT+SA and STRAT+CA condition against the control condition. As we are especially interested in the differential progress of the conditions, the interaction effects with the measurement occasions were also included in the model. To verify whether the level three variation is different for the experimental conditions than it is for the control group, which might arise when the interventions are more effective in some classes than in others, the effects of the STRAT+SA and STRAT+CA dummies were allowed to vary randomly at level three. However, the analyses of these more extensive models did not significantly improve the goodness of fit of the models and revealed no evidence that

there are significant differences between the three research conditions in the variation between classes. Therefore there is no reason to include complex variance at level three for the estimates of the experimental conditions' dummies.

Separate analyses were done for second and fifth graders. The parameters of the models were estimated using the restrictive iterative generalized least squares (RIGLS) estimation procedure of the software MlwiN (Rasbash et al., 1999). Since parsimonious models are preferred, only significant estimates ameliorating the model were retained. In the next section we go into a full description of the final most appropriate models, specifying the efficacy of the experimental interventions in ameliorating second and fifth graders' reading comprehension performance, strategy use, and self-efficacy perceptions toward reading.

## Results

### *Effects on Reading Comprehension Achievement*

Table 6.4 presents the final repeated measures model showing the results for the effects of the interventions on second and fifth graders' reading comprehension achievement after the effects for measurement occasion and explanatory variables such as gender, mother tongue and number of years behind have been taken into account. For second grade achievement scores for reading fluency were also included in the model. Those scores were grand mean centered to facilitate the interpretation of the intercept.

Taken into account all included variables, the intercept of 20.16 in the second-grade model can be interpreted as the overall mean pretest comprehension score across all Dutch-speaking girls in the control group classes, who never stayed down a class and had an average pretest reading fluency score. In the fifth-grade model the intercept is to be interpreted as the overall mean pretest comprehension score across all Dutch-speaking girls in the control group classes, who never stayed down a class. The test battery employed with fifth graders is IRT-modeled. Consequently, an increase from the score at pretest to the scores at the post- and retention test can be interpreted as a progress in reading comprehension. The parameters for the posttest and the retention test measurement occasions for the fifth graders are not significantly different from zero, revealing that these students actually did not experience progress in reading comprehension in the course of the fifth grade and the first term of sixth grade. With regard to second graders' reading comprehension, it has to be mentioned that no suitable IRT-modeled Dutch standardized reading comprehension test battery for second graders was found and could not be developed within the scope of the present study. Moreover, the three tests comprised an unequal number of items. Therefore, the reading comprehension scores for the three measurement occasions are not comparable. Consequently, an increase or decrease from the scores at pretest to the scores at the post- and retention test can not be interpreted as a progress or relapse in reading comprehension.



Table 6.4

*Model Estimates for the Repeated Measures Three-Level Analyses of the Second- and Fifth-Grade Reading Comprehension Achievement Scores*

Parameter	Grade	
	2nd	5th
<i>Fixed</i>		
Intercept	20.16 (0.69)***	42.21 (2.01)***
Post	-7.16 (0.83)***	0.94 (1.01)
Retention	-6.79 (1.05)***	3.03 (1.97)
Boy	-1.57 (0.37)***	-2.64 (1.28)*
Non-native	-5.01 (0.91)***	-7.50 (2.71)**
Post * Non-native	4.26 (0.92)***	
Retention * Non-native	2.22 (1.00)*	
Years behind	-2.79 (0.46)***	-9.71 (1.36)***
Fluency	0.25 (0.02)***	
Post * Fluency	-0.10 (0.02)***	
Retention * Fluency	-0.12 (0.02)***	
STRAT+SA	2.27 (0.94)*	1.10 (2.42)
STRAT+CA	-0.01 (0.87)	-1.88 (2.50)
Post * STRAT+SA	0.69 (1.09)	3.02 (1.28)*
Post * STRAT+CA	2.13 (1.00)*	4.35 (1.32)***
Retention * STRAT+SA	0.07 (1.30)	4.55 (2.59)
Retention * STRAT+CA	1.15 (1.19)	6.59 (2.65)*
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	1.27 (0.82)	8.19 (5.21)
$\sigma_{v0vpost}$	-1.55 (0.87)	
$\sigma^2_{vpost}$	2.12 (1.09)	
$\sigma_{v0vret}$	-1.90 (1.02)	-0.94 (4.08)
$\sigma_{vpostvret}$	2.63 (1.21)*	
$\sigma^2_{vret}$	3.58 (1.57)*	14.13 (6.22)*
Level 2		
$\sigma^2_{\mu0}$	25.37 (1.88)***	220.41 (15.36)***
$\sigma_{\mu0\mu post}$	-16.66 (1.59)***	-70.72 (8.27)***
$\sigma^2_{\mu post}$	24.41 (1.86)***	100.58 (7.19)***
$\sigma_{\mu0\mu ret}$	-15.92 (1.57)***	-74.68 (10.01)***
$\sigma_{\mu post\mu ret}$	15.43 (1.55)***	67.59 (7.19)***
$\sigma^2_{\mu ret}$	23.18 (1.84)***	143.65 (10.84)***
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

As can be seen in Table 6.4, second- and fifth-grade boys perform significantly lower than girls. Effect sizes are 0.25 and 0.17 standard deviation respectively. The absence of any interaction effect with the post- and retention test dummies reveals more specifically that boys perform significantly lower with a fixed arrear at each measurement occasion. A similar but bigger effect was found for being a non-native Dutch speaker in fifth grade ( $ES = 0.49 SD$ ). On the average second-grade non-native speakers also perform significantly less at the pretest ( $ES = 0.68 SD$ ), but they appear to catch up significantly in the course of second and third grade, as is shown by the significant and positive interaction effects between the non-native dummy and the post- and retention test dummies. Further, the number of years second and fifth graders are behind at school has been found inversely proportional to their comprehension

scores, with a fixed impact for all the measurement occasions. Finally, as regards the impact of reading fluency it can be concluded that second graders with higher pretest fluency scores on the average also have higher pretest reading comprehension scores as well. Nevertheless, the relationship between reading fluency and reading comprehension becomes weaker at the post- and retention test.

For second grade the estimated fixed effects of the dummy variables representing the STRAT+SA and STRAT+CA condition are respectively 2.27 and -0.01. These effects are differential intercepts for the experimental conditions compared to the control group, meaning that the average pretest score of Dutch-speaking girls, who never stayed down a class and had an average fluency score is 22.43 in the STRAT+SA and 20.15 in the STRAT+CA condition. The estimates show that students in the STRAT+SA condition on the average have significantly higher scores at the pretest than students in either the control or STRAT+CA condition. The estimates of the interaction effects between the measurement occasions and the STRAT+SA and STRAT+CA condition reveal that the evolutions of the students from pre- to posttest are more favorable in the STRAT+SA and STRAT+CA conditions than in the control condition. More specifically, the estimates correspond with effect sizes of respectively 0.26 and 0.42 standard deviation. However, only the difference between the STRAT+CA and control condition reaches significance ( $\chi^2 = 4.56$ ,  $df = 1$ ,  $p = .033$ ). With regard to the progress over the whole period from pretest to retention test, the effect sizes drop to 0.14 for the STRAT+SA and 0.23 for the STRAT+CA condition. Moreover, the difference between the STRAT+CA and control condition is no longer significant. A pairwise comparison of the experimental conditions reveals no significant difference between the STRAT+SA and STRAT+CA condition at neither the post- nor the retention test.

For fifth grade the estimated main STRAT+SA and STRAT+CA effects indicate that at pretest both experimental groups are not significantly different from the control condition. Moreover, a pairwise comparison of the experimental conditions' estimates neither reveal significant pretest differences. The interaction effects' estimates, however, reveal that the progress from pre- to posttest in both the STRAT+SA ( $\chi^2 = 5.52$ ,  $df = 1$ ,  $p = .019$ ) and the STRAT+CA condition ( $\chi^2 = 10.90$ ,  $df = 1$ ,  $p = .001$ ) is significantly higher than in the control condition. Effect sizes are 0.21 and 0.28 standard deviation respectively. With regard to the progress over the whole period from pretest to retention test only students in the STRAT +CA condition ( $\chi^2 = 6.19$ ,  $df = 1$ ,  $p = .013$ ) experience significant differential learning gains with an effect size of 0.42 standard deviation. The differential slope of the STRAT+SA condition ( $\chi^2 = 3.08$ ,  $df = 1$ ,  $p = .079$ ), with an effect size of 0.28 standard deviation only reaches a marginal significance level. A pairwise comparison of the experimental conditions show no significant difference between the STRAT+SA and STRAT+CA condition's progress from pretest to post- or retention test.

To verify the statistical significance of the separate progress in fifth grade from posttest to retention test an additional analysis was executed with the pretest and retention test contrasted against the posttest measure as reference group. This analysis revealed that neither of the experimental conditions made a post- to retention test

progress that is significantly higher than the control group's progress. A pairwise comparison of the experimental conditions' progress yielded no significant differences either. Similarly, second graders' decrease from post- to retention test was also not significant.

As regards the random part of the second-grade model, a fully multivariate model with regard to the repeated measures appeared to be a strong improvement over the compound symmetry model. As Table 6.4 points out, the analysis shows a marginal significant variance component for the intercept at the class level, a significant intercept variance at the student level, and significant slope variances of both dummies at the student as well as at the class level. This reveals that classes and students within those classes have different initial states and different rates of progress in the course of time. Examining the variance function it can be concluded that the level-three variation between classes regarding their reading comprehension achievement is decreasing from pretest (1.27) to posttest (0.29) and again increasing at retention test (1.04). The level-two variation between students within classes outweighs the differences between classes and also decreases from pretest (25.37) to posttest (16.47), followed by a stabilization at the retention test (16.71).

The random part of the fifth-grade model shows a similar pattern, except for the level-three variation of the posttest dummy. More particularly, it appears that the level-three variation between fifth-grade classes is increasing from pretest (8.19) to retention test (20.44). The level-two variation between fifth-grade students decreases from pretest (220.41) to posttest (179.55), followed by an increase at the retention test (214.70), leading to a variation comparable with the situation at pretest. In both the second- and fifth-grade model, no other estimates were found to vary randomly across classes or students.

### *Effects on Strategy Use*

With regard to students' report of strategy use, similar models were built up. Besides including the dummies representing the measurement occasions and research conditions, the relevance of other explanatory variables were explored as well. The fixed parts of the second- and fifth-grade model in Table 6.5 reveal, however, that second graders' gender was the only explanatory predictor with a significant impact. More specifically, at each measurement occasion second-grade boys report significantly less strategic reading than girls. Since the same questionnaire was used at each measurement, the differences between the three measurement occasions can be interpreted as an evolution. The estimates for the post- and retention test dummies reveal that second-grade control group students report significantly less strategy use at both post- and retention test as compared to the pretest measurement occasion. No significant differences between the post- and retention test were found. Fifth-grade control group students score somewhat higher at pretest. Moreover they also show lower average scores at post and retention test, but in their case the evolution from pre- to post- and retention test was not significant.

Table 6.5

*Model Estimates for the Repeated Measures Three-Level Analyses of the Second- and Fifth-Grade Measures of the Use of Reading Comprehension Strategies*

Parameter	Grade	
	2 <sup>nd</sup>	5 <sup>th</sup>
<i>Fixed</i>		
Intercept	2.83 (0.07)***	2.93 (0.07)***
Post	-0.16 (0.05)**	-0.05 (0.07)
Retention	-0.10 (0.05)*	-0.01 (0.07)
Boy	-0.09 (0.03)**	
STRAT+SA	-0.11 (0.09)	-0.17 (0.09)
STRAT+CA	-0.14 (0.09)	-0.03 (0.10)
Post * STRAT+SA	0.19 (0.07)**	0.07 (0.10)
Post * STRAT+CA	0.32 (0.06)***	0.12 (0.10)
Retention * STRAT+SA	0.14 (0.07)	-0.01 (0.09)
Retention * STRAT+CA	0.27 (0.07)***	0.02 (0.09)
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	0.02 (0.01)*	0.02 (0.10)*
$\sigma_{v0vpost}$		-0.01 (0.01)
$\sigma^2_{vpost}$		0.02 (0.01)*
$\sigma_{v0vret}$		-0.01 (0.01)*
$\sigma_{vpostvret}$		0.02 (0.01)*
$\sigma^2_{vret}$		0.02 (0.01)*
Level 2		
$\sigma^2_{\mu0}$	0.15 (0.02)***	0.15 (0.01)***
$\sigma_{\mu0\mu post}$	-0.13 (0.02)***	-0.06 (0.01)***
$\sigma^2_{\mu post}$	0.27 (0.02)***	0.17 (0.01)***
$\sigma_{\mu0\mu ret}$	-0.14 (0.02)***	-0.06 (0.01)***
$\sigma_{\mu post\mu ret}$	0.17 (0.02)***	0.12 (0.01)***
$\sigma^2_{\mu ret}$	0.29 (0.02)***	0.19 (0.01)***
$\sigma_{\mu0\mu boy}$	0.03 (0.01)*	
$\sigma_{\mu post\mu boy}$	-0.01 (0.02)	
$\sigma_{\mu ret\mu boy}$	-0.02 (0.02)	
$\sigma^2_{\mu boy}$	0.00 (0.00)	
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

The fixed estimates of the dummy variables representing the experimental conditions indicate no significant pretest differences between the three research conditions, neither for second grade, nor for fifth grade. Contrary to control group students both second-grade experimental conditions report more frequent use of reading strategies at post- and retention test. For the posttest the difference with the control group is significant for both the STRAT+SA ( $\chi^2 = 7.22$ ,  $df = 1$ ,  $p = .007$ ) and the STRAT+CA condition ( $\chi^2 = 26.09$ ,  $df = 1$ ,  $p = .000$ ). Effect sizes are 0.42 and 0.71 standard deviation respectively. But only for the STRAT+CA condition the difference with the control group students remains significant at the retention test ( $\chi^2 = 16.38$ ,  $df = 1$ ,  $p = .000$ ) ( $ES = 0.62$   $SD$ ). However, the difference between the STRAT+SA and control condition ( $\chi^2 = 3.39$ ,  $df = 1$ ,  $p = .066$ ) still reaches marginal significance ( $ES = 0.31$   $SD$ ). A pairwise comparison of the experimental conditions reveals significant

differences between the second-grade STRAT+SA and STRAT+CA condition as well. More specifically, a significantly larger increase in the use of reading strategies is reported by second graders in the STRAT+CA condition, both at post- ( $\chi^2 = 4.40$ ,  $df = 1$ ,  $p = .036$ ) and retention test ( $\chi^2 = 3.88$ ,  $df = 1$ ,  $p = .049$ ), respectively with effect sizes of 0.29 and 0.31 standard deviation. With regard to the fifth graders' evolution from pre- to post- and retention test, no significant differences were found between the three research conditions.

As regards the random part of the second-grade model, Table 6.5 indicates significant level-two variation for the estimates of the measurement occasion and the gender estimate. More specifically, the level-two variation increases slightly, but significantly from pretest (0.15 for girls) to posttest (0.16 for girls) and retention test (0.16 for girls). Moreover, with regard to second graders' gender it can be concluded that, in addition to the finding that boys report on average significantly less use of reading strategies, the variation between students is significantly larger for boys (e.g., 0.20 at pretest) than for girls (e.g., 0.15 at pretest).

With regard to fifth graders, the random part of the model reveals significant variation at level two and three for the measurement occasion estimates. Examining the variance function it can be concluded that the variation between classes is slightly increasing from pretest (0.02) to posttest (0.03) and again decreasing at retention test (0.02). The level-two variation between students within classes increases from pretest (0.15) to post- (0.21) and retention test (0.22).

### *Effects on Self-Efficacy Thoughts and Attributions*

As regards students' thoughts about reading proficiency and success or failure attributions, statistically significant effects for second graders were only found for thoughts relating to success attributions and positive self-efficacy perceptions towards reading. For fifth graders, on the other hand, significant effects were found for thoughts relating to failure attributions and negative self-efficacy perceptions. The parameter estimates are presented in Table 6.6.

The fixed parts of the models show that besides the dummies representing the measurement occasions and research conditions, and the interaction effects between those, second graders' gender and fifth graders' mother tongue yielded significant estimates and improvements of the models. More specifically, at each measurement occasion second-grade boys report significantly less thoughts with regard to success attributions and positive self-efficacy perceptions than girls. As concerns the fifth graders, at each measurement occasion non-native students report significantly more thoughts with regard to failure attributions and negative self-efficacy perceptions than Dutch-speaking fifth graders. Taken into account the included explanatory variables, the second-grade model's intercept of 3.45 can be read as the mean pretest score with regard to positive thoughts about one's own reading abilities and associated success attributions across all girls being part of the control group classes. On the other hand, the intercept of 1.88 in the fifth-grade model represents the mean pretest score with

regard to negative thoughts about one's own reading abilities and associated failure attributions across all native speakers being part of the control group classes.

From pre- to post- and retention test second-grade control group scores drop significantly. No significant control group differences were found between the post- and retention test. The estimates of the dummy variables representing the STRAT+SA ( $\chi^2 = 4.55$ ,  $df = 1$ ,  $p = .032$ ) and STRAT+CA condition ( $\chi^2 = 8.17$ ,  $df = 1$ ,  $p = .004$ ) show that second graders' in both experimental groups differ significantly in their pretest scores. Effect sizes are 0.40 and 0.49 standard deviation respectively. More specifically, they report significantly less thoughts with regard to success attributions and positive self-efficacy perceptions than their control group peers. A pairwise comparison of both experimental conditions' estimates reveals no significant pretest differences. As concerns the experimental conditions' differential change from pre- to post- and retention test with regard to the control group's evolution, Table 6.6 reveals that both the STRAT+SA ( $\chi^2 = 5.80$ ,  $df = 1$ ,  $p = .016$ ) and STRAT+CA ( $\chi^2 = 11.39$ ,  $df = 1$ ,  $p = .001$ ) conditions' change from pre- to posttest is significantly higher than the control group's evolution. Effect sizes are 0.44 and 0.56 standard deviation respectively. Contrary to the control group students reporting significantly less thoughts with regard to success attributions and positive self-efficacy perceptions at posttest, students in the experimental conditions' reports remain more or less the same. With regard to the progress over the whole period from pretest to retention test, however, the significant difference with respect to the control group disappeared for both the STRAT+SA and STRAT+CA condition. Neither at posttest, nor at retention test pairwise comparisons of the experimental conditions reveal significant differences between the STRAT+SA and STRAT+CA condition.

With regard to the fifth-grade model, Table 6.6 reveals that there is no significant control group evolution from pre- to post- and retention test. Further, the estimates of the STRAT+SA and STRAT+CA dummies reveal no significant pretest differences between the three conditions. The interaction effects between the measurement occasions and the conditions, however, indicate that by the end of the school year children engaged in same-age tutoring activities are significantly less occupied with failure attributions and negative self-efficacy related thoughts compared to students in the control ( $\chi^2 = 11.95$ ,  $df = 1$ ,  $p = .001$ ) and students in the STRAT+CA condition ( $\chi^2 = 7.68$ ,  $df = 1$ ,  $p = .006$ ). At the retention test, the difference with the control group became only marginal significant ( $\chi^2 = 3.62$ ,  $df = 1$ ,  $p = .057$ ). Remarkably, at the retention test a marginal significant effect was observed for the STRAT+CA condition as well ( $\chi^2 = 3.03$ ,  $df = 1$ ,  $p = .082$ ). Consequently, the significant difference between the STRAT+SA and STRAT+CA condition disappeared at retention test.

As regards the random parts of the models, Table 6.6 indicates significant level-two variation for the estimates of the measurement occasion for both the second- and fifth-grade model. More specifically, the level-two variation in second graders' reports of success attributions and positive self-efficacy perceptions toward reading increases from pretest (0.43) to posttest (0.57) and decreases again to the retention test (0.49). The variation between fifth-grade classes in students' reports of failure attributions and

negative self-efficacy perceptions follow the same trend from pretest (0.26) to post- (0.30) and retention test (0.28). No other second- or fifth-grade estimates were found to vary randomly across second- of fifth-grade classes or students.

Table 6.6

*Model Estimates for the Three-Level Analyses of the Second-Grade Measures of Success Attributions and Positive Self-efficacy Perceptions and Fifth-Grade Measures of Failure Attributions and Negative Self-efficacy Perceptions*

Parameter	Grade	
	2nd	5th
<i>Fixed</i>		
Intercept	3.45 (0.09)***	1.88 (0.05)***
Post	-0.33 (0.09)***	0.04 (0.05)
Retention	-0.31 (0.09)***	-0.05 (0.05)
Boy	-0.13 (0.05)*	
Non-native		0.25 (0.09)**
STRAT+SA	-0.27 (0.13)*	0.10 (0.06)
STRAT+CA	-0.34 (0.12)**	0.03 (0.07)
Post * STRAT+SA	0.31 (0.13)*	-0.22 (0.06)***
Post * STRAT+CA	0.39 (0.12)***	-0.06 (0.06)
Retention * STRAT+SA	0.11 (0.12)	-0.12 (0.06)
Retention * STRAT+CA	0.12 (0.11)	-0.11 (0.06)
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	0.03 (0.01)*	0.00 (0.00)
Level 2		
$\sigma^2_{\mu0}$	0.43 (0.03)***	0.26 (0.02)***
$\sigma_{\mu0\mu post}$	-0.36 (0.04)***	-0.11 (0.01)***
$\sigma^2_{\mu post}$	0.86 (0.07)***	0.25 (0.02)***
$\sigma_{\mu0\mu ret}$	-0.35 (0.04)***	-0.11 (0.01)***
$\sigma_{\mu post\mu ret}$	0.49 (0.05)***	0.14 (0.01)***
$\sigma^2_{\mu ret}$	0.76 (0.06)***	0.24 (0.02)***
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

## Discussion

The main aim of the study reported in the present article was the evaluation of two sets of instructional interventions in real second- and fifth-grade classrooms as tools for ameliorating students' reading comprehension achievement, use of reading comprehension strategies, and self-efficacy perceptions towards reading. More particularly, the interventions comprised two innovative cornerstones, namely explicit instruction in reading comprehension strategies and opportunities to practice strategic reading by stimulating interaction between peers about texts in either cross-age or reciprocal same-age peer tutoring dyads. A pretest posttest retention test design including two experimental conditions and a comparable control condition was used to investigate the effectiveness of the interventions and the durability of the potential effects. Multilevel linear regression models were applied to analyze the data of standardized reading comprehension tests and questionnaires with regard to the use of

reading strategies and thoughts about one's own reading proficiency and accompanying success and failure attributions.

### *Effects on Reading Comprehension Achievement*

The analysis with regard to second graders' reading comprehension achievement revealed that students who practiced explicitly taught reading strategies in cross-age peer tutoring dyads outperformed their control group peers at the end of the school year with an effect size of 0.42 standard deviation. In third grade, however, six months after the end of the experimental interventions, the significant effect disappeared. No significant effects were found for second graders practicing the reading strategies in reciprocal same-age dyads. These results parallel the outcomes of an earlier study (Van Keer & Verhaeghe, 2002) and corroborates the efficacy for second-grade students of explicit reading strategies instruction followed by practice in peer tutoring activities led by a fifth-grade tutor.

Although second graders' reading comprehension skills may be graded up through cross-age tutoring, the retention test data leads us to suspect that only by continuing the interventions, long-lasting treatment effects can be maintained. Apparently, practicing reading strategies in second-grade same-age pairs, as developed and administered in the study, is aiming to high. Additional research is necessary to explore the causes of the disappointing STRAT+SA results and to verify whether the preparatory tutor training was not appropriate or unsatisfactory for the second-grade age group, whether the selected reading strategies were too difficult to practice independently in same-age dyads, or whether same-age peer tutoring is not an appropriate instructional technique for this age group. Moreover, subsequent research should investigate whether sustained second-grade STRAT+CA effects can be attained by pursuing the interventions.

Contrary to second-grade results, fifth graders in both the STRAT+SA and STRAT+CA condition made a significantly higher progress than the control group students by the end of the school year, with effect sizes of respectively 0.21 and 0.28 standard deviation. However, only for the students acting as tutor for second graders the significant effect retained six months after the end of the interventions. Interestingly, these students' extra learning gain even increased at this point in time, reaching an effect size of 0.42 standard deviation. The endurance and magnitude of this STRAT+CA effect underscores that practicing the reading strategies as tutor in cross-age dyads produces the most favorable and durable results for fifth graders. However, the results indicate that the significant differential progress is especially situated in the period from pre- to posttest, which is exactly the phase during which the interventions were implemented. Again, the findings of the present study accord with the results of the previous intervention study of Van Keer and Verhaeghe (2002). The magnitude of the effect sizes is however less pronounced, but this is probably attributable to the significant lower pretest scores of the STRAT+CA condition in the first study.



A possible explanation for the significant STRAT+SA and STRAT+CA posttest results can be found in the tutoring task aimed at practicing strategic reading and in the responsibility fifth graders accepted for supporting other students. Taking on the tutor role involves that students are challenged to consider the subject fully from different perspectives, to engage in active monitoring to identify and correct errors, to reorganize and clarify their own knowledge and understandings, and to elaborate on information in their explanations in order to provide help that can be useful to the tutee (Fuchs & Fuchs, 2000; Wittrock, 1990 in Terwel, Gillies, van den Eeden, & Hoek, 1999). Moreover, specifically with respect to the mastery of reading strategies, the fact not having to read themselves may create more opportunities for tutors to give full attention to the metacognitive skills involved in monitoring and regulating the reading process. Consequently, it may be more appropriate to practice the metacognitive skills involved by monitoring and regulating someone else's reading process compared to one's own reading process.

As concerns long-term results, tutoring a younger student with a clearly lower reading level seems especially favorable. Some non-systematic observations of same-age and cross-age reading dyads indicate that the absence of a significant long-term improvement of reading comprehension achievement for the STRAT+SA condition is possibly due to a gradual fading away of the reciprocal tutor and tutee roles, less academic responding, a decrease in time-on-task, and less highly occupied activity in the same-age dyads. Of course, more systematic qualitatively oriented analyses of the quality and nature of the interactions in the reading dyads are requisite to test these assumptions.

### *Effects on Strategy Use*

With regard to students' reports of the use of reading comprehension strategies, analyses revealed only significant effects for second graders. However, in this respect, one has to take into account that fifth graders already scored somewhat higher on the questionnaire at the pretest. More specifically, the results show that compared to their control group counterparts second graders practicing the reading strategies in same-age and cross-age peer tutoring activities reported a significantly higher increase in strategy use by the end of the school year, with effect sizes of respectively 0.42 and 0.71 standard deviation. Six months after the end of the intervention, however, only the significant difference of the STRAT+CA students endured, with an effect size of 0.62 standard deviation. Interestingly, both at post- and retention test, second graders from cross-age dyads reported a significantly higher growth in their reports of strategy use than students in the same-age teams.

These second-grade results suggest that by practicing the application of reading strategies in reciprocal same-age and cross-age reading dyads second graders become more aware of the benefits of using reading strategies during the reading process to enhance the likeliness of comprehension. Moreover, fifth graders functioning as tutor for second graders even seem more effective in accomplishing this awareness. Of course, subsequent research is necessary to confirm the present second-grade results and to try to elucidate the failure to show progress in fifth graders reports of strategy

use. In this respect the use of thinking-aloud protocols could be an interesting supplement to the questionnaire used to shed light on students' reading behavior, application of reading strategies, and metacognitive regulation of the reading process.

### *Effects on Self-Efficacy Thoughts and Attributions*

With regard to students' thoughts about their reading proficiency and accompanying success and failure attributions, no concurrent results were found for second and fifth graders. For second graders we did not find any effects on being preoccupied with negative self-efficacy related thoughts. The analyses, however, revealed that contrary to the control group students reporting significantly less thoughts with regard to success attributions and positive self-efficacy perceptions by the end of the school year, the experimental students reported more or less the same amount of positive thoughts. Nevertheless, six months after the end of the interventions, the experimental students' evolution did no longer vary significantly from the control group's change. In interpreting the pre- to posttest evolution, one has to take into account, however, that at pretest second graders' in both experimental groups reported significantly less thoughts with regard to positive self-efficacy perceptions and accompanying success attributions than their control group peers. The fact that both the STRAT+SA and STRAT+CA condition are not equivalent on the positive self-efficacy related thoughts at pretest imports a serious threat to the internal validity of this part of the study. Therefore, study replications with commensurable research conditions at pretest are needed to verify the positive results of the present research. One has also to bear in mind that a decrease in reported positive self-efficacy thoughts does not necessarily mean that students' self-esteem grow worse. Strictly spoken it merely indicates being less preoccupied with thoughts about that.

As regards fifth graders' preoccupation with thoughts relating to self-efficacy in reading and attributions towards reading, statistically significant effects were found for thoughts relating to failure attributions and negative self-efficacy perceptions. More specifically, by the end of the school year the analyses indicated a significant decrease in the reports of failure attributions and negative self-efficacy related thoughts for children engaged in same-age tutoring activities as compared to students in the control and students in the STRAT+CA condition. As it can be assumed that high scores for thoughts relating to failure attributions and negative self-efficacy perceptions indicate the presence of low self-esteem with regard to reading ability, these results indicate that opportunities to alternate between the tutee and tutor role in reciprocal same-age dyads have a positive impact on fifth graders' self-esteem and perceived competence. Contrary to the STRAT+SA condition's results, the findings with respect to the STRAT+CA condition are not unambiguous. By the end of the school year the decline in the reports of negative self-efficacy related thoughts is significantly smaller than that of their peers in the STRAT+SA condition. However, six months later the differential change from pretest to retention test with respect to the control group becomes marginally significant and the significant difference as compared to the STRAT+SA condition disappears. Surprisingly, the results of both experimental

conditions are exactly opposite to the findings in the previous intervention study of Van Keer and Verhaeghe (2002). In that study it was found that on both post- and retention test fifth graders in the STRAT+CA condition reported significantly less negative thoughts related to their reading proficiency. Clearly, additional research is necessary to shed light on these contrasting outcomes. Again, thinking-aloud protocols could be an interesting supplement to assess students' occupation with self-efficacy perceptions and accompanying attributions while reading.

### *Conclusion*

In accordance with previous research referred to above, the present study generally corroborates the first hypothesis with regard to students' reading comprehension performance on standardized reading tests. More specifically, the study documents the feasibility of fostering reading comprehension achievement in second graders by supplying them with explicit instruction in reading strategies, followed by practice in peer tutoring activities led by a fifth-grade tutor. The same applies for fifth graders, receiving explicit reading strategies instruction and opportunities to practice in cross-age or reciprocal same-age peer tutoring sessions. However, only for fifth graders acting as tutor for second graders in cross-age dyads, long-term effects can be documented.

As to the second and third hypothesis, it was assumed that practicing the application of reading strategies in cross-age dyads would be more effective in fostering second and fifth graders' reading comprehension achievement than practice in reciprocal same-age dyads. Since in second grade the STRAT+CA condition significantly outperformed the control condition at posttest, whereas the STRAT+SA condition was not significantly different from this reference group, the results point in this direction. The same reasoning applies to fifth graders' retention test results. Notwithstanding these indications, no mutually significant post- or retention test differences between the tutoring conditions could be recorded.

As concerns students' strategy use, the study only confirms the hypotheses with regard to second graders. More specifically, both the STRAT+SA and STRAT+CA interventions enhanced second graders' posttest reports of strategy use significantly more than traditional reading comprehension instruction. Moreover, significant long-term effects from pretest to retention test were also found for the STRAT+CA condition. In addition, the growth in the report of strategy use was significantly larger for second graders functioning as tutees in cross-age tutoring activities than for their peers alternating between the tutor and tutee role in reciprocal same-age activities. For fifth grade neither the first nor the third hypothesis could be confirmed.

With regard to the effects of peer tutoring activities on students' preoccupation during reading with self-efficacy related thoughts and attributions, no generally applicable conclusions can be recorded. For second graders only the first hypothesis is confirmed with regard to students posttest reports of positive self-efficacy perceptions toward reading and accompanying success attributions. More specifically, it is revealed that

contrary to the control group students reporting significantly less positive self-efficacy related thoughts by the end of the school year, the reports of students in the experimental conditions remained stable. However, no significant mutual differences were found between the experimental conditions.

With regard to fifth graders, only effects with regard to negative thoughts were found. The decrease in such thoughts indicates that students became more confident in their own reading competence and were less preoccupied with doubts. This was particularly true at the end of the school year for fifth graders who had been engaged in same-age tutoring activities. So, for this condition and measurement occasion the first hypothesis can be confirmed. Since at the end of the school year children engaged in reciprocal same-age activities were also significantly less occupied with negative self-efficacy related thoughts as compared to students in the cross-age peer tutoring condition, the third hypothesis is corroborated for the posttest measurement occasion in precisely the opposite way. This does however no longer apply to the retention test results.

To conclude, two limitations of the present study should be reported. A first restriction can be mentioned on methodological grounds. Due to the complexity of the interventions, the design does not allow to draw conclusions about the relative contribution of the different constituent components of the intervention, nor was this our purpose. Taken into account previous intervention studies referred to above, it can be assumed that different elements make a valuable contribution to the positive effects, notably (1) the content of teaching and learning (i.e., the selected reading strategies), (2) the instructional approach focusing on explicit teacher explanations, modeling, and a gradual transfer from teacher to student regulation, (3) the regular opportunities to practice strategic reading through structured assignments in well-prepared peer tutoring dyads, and (4) the support to the teachers including an elaborated manual with lesson scenarios and all necessary teacher and student materials, and in-service training and assistance. An exploration of the essential and perhaps dispensable components of the interventions requires a profound component analysis. However, Brown et al. (1996) argue that the approach of the present research is appropriate and defensible when the interest is in evaluating the efficacy of a multicomponential instructional package.

A second problematic aspect of the study has to do with the fact that the control group data were adopted from a previous study (Van Keer & Verhaeghe, 2002). Due to the scope and the means of the present study, collecting new control group data was not achievable. However, in this respect one has to take into account that the experimental and control group data were gathered during two consecutive school years. Moreover, the data collection ran completely parallel at each measurement occasion.

Notwithstanding the limitations, the findings of the present study carry significant implications for reading comprehension instruction in general education classes. With regard to reading comprehension achievement, statistically significant and practically important effects particularly indicated the effectiveness of explicit reading strategies instruction plus cross-age peer tutoring activities for both second and fifth graders and suggest that these interventions are a feasible tool to enhance elementary students' reading comprehension achievement. However, future research must explore methods

of maximizing peer potential. Especially with respect to students' strategy use and thoughts relating to self-efficacy in reading subsequent more in-depth research is needed to clear the lack of univocal results.

## References

- Aarnoutse, C. A. J. (1995). *Probleemidentificatie en aanzet voor een actieplan taal* [Problem identification and initiative towards an action plan for language]. Nijmegen, The Netherlands: Vakgroep Onderwijskunde.
- Aarnoutse, C. A. J., & van de Wouw, J. (1990). *Wie dit leest. Handleiding* [Who reads this. Manual]. Tilburg, The Netherlands: Zwijsen.
- Aarnoutse, C. A. J., & Van Leeuwe, J. F. J. (1998). Relation between reading comprehension, vocabulary, reading pleasure, and reading frequency. *Educational Research and Evaluation*, 4, 143-166.
- Aarnoutse, C. A. J., & Van Leeuwe, J. F. J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation*, 6, 251-278.
- Aarnoutse, C. A. J., & Weterings, A. C. E. M. (1995). Onderwijs in begrijpend lezen [Education in reading comprehension]. *Pedagogische Studiën*, 72, 82-101.
- Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.
- Alvermann, D. E. (2000). Classroom talk about texts: Is it dear, cheap or a bargain at any price? In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 170-192). New York: Teachers College Press.
- Ames, C. (1984). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Educational Psychology*, 76, 478-487.
- Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353-394). New York: Longman.
- Bentz, J. L., & Fuchs, L. S. (1996). Improving peers' helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly*, 19, 202-215.
- Brand-Gruwel, S., Aarnoutse, C. A. J., & Van Den Bos, K. P. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and instruction*, 8, 63-81.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.
- Brus, B. Th. (1969). *Eén-minuut-test* [One minute test]. Nijmegen, The Netherlands: Berkhout.
- Chard, D. J., Simmons, D. C., & Kameenui, E. J. (1998). Word recognition: Research bases. In D. C. Simmons, & E. J. Kameenui, (Eds.), *What reading research tells us about children with diverse learning needs. Bases and basics* (pp. 141-167). Mahwah, NJ: Lawrence Erlbaum Associates.

- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.
- De Corte, E., Verschaffel, L., & Van De Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Educational Psychology*, 71, 531-559.
- Dole, J. A. (2000). Explicit and implicit instruction in comprehension. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 52-69). New York: Teachers College Press.
- Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61, 239-264.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliff, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putnam, J., & Bassari, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22, 347-368.
- Durkin, D. (1978-1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14, 481-533.
- Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on self-concept and mathematics achievement. *Journal of Educational Psychology*, 87, 272-281.
- Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology*, 84, 331-339.
- Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *The Journal of School Psychology*, 28, 309-323.
- Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81, 173-177.
- Fuchs, L. S., & Fuchs, D. (2000). Building student capacity to work productively during peer-assisted reading activities. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 95-115). New York: Teachers College Press.
- Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, C. L. (1994). The nature of students interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75-103.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K. & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 97, 223-249.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [Step out! Manual]. Utrecht, The Netherlands: Sardes.

- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Greenwood, C. R., Carta, J. C., & Hall, R. V. (1988). The use of peer tutoring strategies in classroom management and educational instruction. *School Psychology Review*, 17, 258-275.
- Greenwood, C. R., & Delquadri, J. C. (1995). Classwide peer tutoring and the prevention of school failure. *Preventing School Failure*, 39, 21-25.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.
- Greenwood, C. R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The classwide peer tutoring program: Implementation factors moderating students achievement. *Journal of Applied Behavior Analysis*, 25, 101-116.
- Greenwood, C. R., Terry, B., Utley, C. A., Montagna, D., & Walker, D. (1993). Achievement, placement and services: Middle school benefits of classwide peer tutoring used at the elementary school. *School Psychology Review*, 22, 497-516.
- Haller, E. P., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher*, 17, 5-8.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 33-68). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2, 127-160.
- Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: verbal versus visual strategies. *Journal of Educational Psychology*, 92, 772-782.
- King-Sears, M. E., & Bradley, D. F. (1995). Classwide peer tutoring. *Preventing School Failure*, 40, 29-36.
- Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, 96, 275-293.
- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, 99, 3-21.
- Krull, J. L. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 428-444.
- Lambiotte, J. G., Dansereau, D. F., O'Donnell, A. M., Young, M., D., Skaggs, L. P., Hall, R. H., & Rocklin, T. R. (1987). Manipulating cooperative scripts for teaching and learning. *Journal of Educational Psychology*, 79, 424-430.
- Lories, G., Dardenne, B., & Yzerbyt, V. Y. (1998). From social cognition to metacognition. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition. Cognitive and social dimensions* (pp. 1-15). London: Sage.
- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.
- Mathes, P. G., Torgesen, J. K., Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted

- instruction in phonological awareness. *American Educational Research Journal*, 38, 371-410.
- Marsh, H. W. (1984). Relations among dimensions of self-attribution, dimensions of self-concepts and academic achievements. *Journal of Educational Psychology*, 76, 1291-1308.
- Marsh, H. W., Cairns, L., Relich, J., Barnes, J., & Debus, R. L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. *Journal of Educational Psychology*, 76, 3-32.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Palincsar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. *Remedial and Special Education*, 12, 43-53.
- Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition and motivation. *Developmental Review*, 6, 25-56.
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research: Vol. 2* (p. 609-640). New York: Longman.
- Plewis, I., & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation*, 4, 13-26.
- Pressley, M. (2000). Comprehension instruction in elementary school: A quarter-century of research progress. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 32-51). New York: Teachers College Press.
- Pressley, M., & Afflerbach, P. (1995). *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading*. Hillsdale, NJ: Erlbaum.
- Pressley, M., & Allington, R. (1999). What should reading instructional research be the research of? *Issues in Education*, 5, 1-35.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92, 513-555.
- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal*, 90, 3-32.
- Pressley, M., Wharton-McDonald, R., Hampston, J. M., & Echevarria, M. (1998). The nature of literacy instruction in ten grade-4/5 classrooms in upstate New York. *Scientific Studies of Reading*, 2, 159-191.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, 64, 479-530.
- Roswal, G. M., Mims, A. A., Evans, M. D., Smith, B., Young, M., Burch, M., Croce, R., Horvat, M. A., & Block, M. (1995). Effects of collaborative peer tutoring on urban seventh graders. *Journal of Educational Research*, 88, 275-279.



- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced modeling*. London: Sage.
- Staphorsius, G., & Krom, R. (1996). *Toetsen Begrijpend Lezen* [Reading Comprehension Tests]. Arnhem, The Netherlands: Cito.
- Terwel, J., Gillies, R. M., van den Eeden, P., & Hoek, D. (1999). *Cooperative learning processes of students: A longitudinal multilevel perspective*. Paper presented at the American Educational Research Association Conference, Montreal, Canada (19-23 April, 1999).
- Topping, K. J. (1996) The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321-345.
- Utley, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children*, 29, 69-92.
- Van Den Broek, P., & Kremer, K. E. (2000). The mind in action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 1-31). New York: Teachers College Press.
- Van Keer, H. (in press). *Een boek voor twee. Verwerving van strategieën voor begrijpend lezen via peer tutoring*. [A book for two. Acquisition of reading comprehension strategies through peer tutoring.] Leuven, Belgium: Garant.
- Van Keer, H. (2002). *Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring*. Manuscript submitted for publication.
- Van Keer, H., & Verhaeghe, J. P. (2002). *Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions*. Manuscript submitted for publication.
- Verhoeven, L. (1993). *Lezen met begrip 1* [Reading with comprehension 1]. Arnhem, The Netherlands: Cito.
- Walraven, M. (1995). *Instructie in leesstrategieën. Problemen met begrijpend lezen en het effect van instructie aan zwakke lezers* [Instruction in reading strategies. Difficulties with reading comprehension and the effect of instruction for poor readers]. Amsterdam: Paedologisch Instituut.
- Weisberg, R. (1988). 1980s: A change in focus of reading comprehension research: A review of reading/learning disabilities research based on an interactive model of reading. *Learning Disability Quarterly*, 11, 149-159.
- Weterings, A. C. E. M., & Aarnoutse, C. A. J. (1986). De praktijk van het onderwijs in begrijpend lezen [Instruction in reading comprehension in practice]. *Pedagogische Studiën*, 63, 387-400.
- Wiersma, W. (2000). *Research Methods in Education. An Introduction*. Boston: Allyn and Bacon.

## Chapter 7

### General discussion

In the present dissertation an effort was made to narrow the yawning gap between prevailing instructional practice and research evidence in the field of reading comprehension instruction. More specifically, the main purpose of the dissertation was to design, implement, and evaluate instructional innovations, aimed at substantially revising reading comprehension instruction in elementary schools and making it more effective. Based on the existing research evidence indicating explicit reading strategies instruction and regular peer-led interaction about texts as significant strategies to foster students' reading comprehension skills, the innovative approach developed within the framework of the dissertation blended research-based practices from both research fields. More specifically, the innovations comprised two innovative cornerstones, namely explicit instruction in six relevant reading strategies and opportunities to practice strategic reading by stimulating interaction between peers about texts in either cross-age or reciprocal same-age peer tutoring dyads. During two successive school years, two similar large-scale and quasi-experimental studies were conducted to inquire into the effectiveness and the durability of the experimental treatments on regular second and fifth graders' reading skills and social and emotional development. Chapter 2, 3, 4, and 6 presented the results of both studies in detail.

By linking reading strategies instruction with peer tutoring, the studies built on former research. Moreover the studies extended prior research by (a) designing a large-scale investigation and sampling a relatively large number of participants from 25 different schools throughout Flanders; (b) implementing the experimental treatments in ecologically valid settings by supporting teachers to introduce the innovative approach as part of the overall curriculum, in the regular classroom context and with the participation of all students during an entire school year; (c) including long-term retention assessments; (d) using standardized reading comprehension tests instead of experimenter-designed tests; (e) applying multilevel modeling to take into account the hierarchical nesting of students in classes and the quasi-experimental design of the studies; (f) enabling an explicit exploration of the surplus value of peer tutoring in comparison with teacher-led practice of strategic reading; and (g) comparing the differential effects of cross-age and reciprocal same-age peer tutoring within the same study. The latter element is unique with respect to other peer tutoring studies, that merely focused on studying the effectiveness of only one tutoring variant. So far, there is only indirect evidence that cross-age tutoring is more effective than same-age tutoring, for the meta-analysis of Cohen, Kulik, and Kulik (1982) reveals larger effect sizes for both tutors' and tutees' achievement in cross-age tutoring programs. However, no statistically significant differences between the variants were revealed.

In addition, the efficacy of two teacher training strategies to disseminate the innovative approaches into regular instructional practice was compared. More specifically, a year-round intensive coaching of the teachers was equated to a more restricted 13-hours in-service teacher training course. Criteria to test and compare the efficacy were teachers' experiences with the attended training course and students' progress in reading skills

and social and emotional development. The results of this comparison of training programs were reported in Chapter 5.

This final chapter reviews the results of both studies. More specifically, since the effectiveness of the interventions on students' reading comprehension achievement, reading strategy use and self-efficacy perceptions towards reading was explored twice using different analysis techniques, the general discussion of the research findings and their links with the postulated hypotheses is preceded by an overview assembling the results of both successive studies on student outcomes. Furthermore, some limitations of the reported studies are outlined and a set of suggestions for future research are presented. Finally, the chapter ends with a brief conclusion.

### **Overview of the Results with regard to Student Outcomes**

Two longitudinal studies were set up to address the first three research questions with regard to the differential effectiveness and the durability of explicit reading strategies instruction, followed by practice in either teacher-led whole-class activities, reciprocal same-age, or cross-age peer tutoring activities, on second and fifth graders' reading skills and social and emotional development. More specifically, the studies were conducted in the context of a relatively large number of naturally constituted elementary school classes to provide a ecologically valid setting for the implementation of the interventions. Pretest posttest retention test designs including experimental conditions and a commensurable control condition were used and multilevel linear regression models were applied to analyze the collected student data.

As already stated expressly in the general introduction of the dissertation and as appeared from the successive chapters, two distinct analysis techniques were applied to the data. Initially, student data were analyzed by applying separate hierarchical regression analyses on students' post- and retention test data. More specifically, two levels were distinguished, namely students nested within classes. Taken into account the specialized literature (Goldstein, 1995; Jones & Duncan, 1998; Paterson & Goldstein, 1991; Rasbash et al., 1999; Snijders & Bosker, 1999), the analysis subsequently switched to a technique better adjusted to the longitudinal design of the study with three waves of student data. More particularly, three-level repeated measures models were used, where the lowest-level units are the measurement occasions, clustered within students, who in turn are nested within classes. This evolution in applied data analysis techniques reflects our more profound exploration of the possibilities and applications of hierarchical regression analysis. However, the change of techniques also entails that the comparison of the research results from the first and second study, respectively presented in Chapter 2 and 6 can not be drawn correctly. Therefore the data collected within the framework of the first study were reanalyzed by means of hierarchical repeated measures models. Hereafter, the results of these analyses are outlined shortly. More specifically, summarizing tables assemble the results of the successive studies. The construction and interpretation of the revised models are not discussed elaborately, but run completely parallel to the models described from Chapter 3 onwards.

Just as in the preceding chapters, we respectively use the abbreviations STRAT, STRAT+SA and STRAT+CA to represent the experimental conditions characterized by explicit reading strategies instruction, followed by practicing the strategies in teacher-led whole-class settings, in reciprocal same-age, or in cross-age peer tutoring dyads.

### *Effects on Reading Comprehension Achievement*

Table 7.1 and 7.2 present the final repeated measures models for both studies, respectively revealing the effectiveness of the experimental interventions on second and fifth graders' reading comprehension achievement. The second- and fifth-grade model revealing the findings of the second study are identical to the models reported in Chapter 6. Likewise, the fifth-grade model for the first study was already reported in Chapter 3. As concerns the second graders' reading comprehension achievement in the first study a new model, constructed by means of a three-level repeated measures analysis, was estimated.

With regard to the interpretation of the estimates, it should be mentioned that the test battery employed with fifth graders is IRT-modeled. Consequently, an increase from the score at pretest to the scores at the post- and retention test can be interpreted as a progress in reading comprehension. With regard to second graders, however, no suitable IRT-modeled Dutch standardized reading comprehension test battery for second graders was found and could not be developed within the scope of the studies. Moreover, the three tests comprised an unequal number of items. Therefore, the reading comprehension scores for the three measurement occasions are not comparable. Consequently, an increase or decrease from the score at pretest to the scores at the post- and retention test can not be interpreted as a progress or relapse in reading comprehension, but solely as the reproduction of the differences between the average scores. The interaction effects between the dummies representing the post- and retention test measurement occasion and the dummies representing the experimental STRAT, STRAT+SA and STRAT+CA conditions represent to what extent the experimental conditions' change from pre- to post- and retention test differs from the control group's evolution. The coefficients for these interactions can be compared and interpreted correctly as the extent to which the evolutions in the corresponding experimental groups are more favorable or more disadvantageous than the one in the control group.

Table 7.1 reports significant pretest differences in the second study after taking into account students' gender, mother tongue, number of years behind at school, and reading fluency scores. More specifically, it was found that students in the STRAT+SA condition started second grade with significantly higher reading comprehension achievement scores than students in either the control or STRAT+CA condition. No pretest differences were found in the first study.

With regard to fifth graders, the coefficients in Table 7.2 reveal that, controlled for students' gender, mother tongue, and number of years behind at school, fifth graders in

the first study's STRAT+CA condition performed significantly lower at pretest than the control group students. No pretest differences were found in the second study.

Table 7.1

*Model Estimates for the Repeated Measures Three-Level Analyses of the Second-Grade Reading Comprehension Achievement Scores for Both Studies*

Parameter	Study	
	1 <sup>st</sup>	2 <sup>nd</sup>
<i>Fixed</i>		
Intercept	19.10 (0.68)***	20.16 (0.69)***
Posttest (control group)	0.08 (0.71)	-7.16 (0.83)***
Retention test (control group)	1.51 (0.81)	-6.79 (1.05)***
Boy	-1.04 (0.38)**	-1.57 (0.37)***
Non-native	-5.18 (0.91)***	-5.01 (0.91)***
Posttest * Non-native	2.20 (0.83)**	4.26 (0.92)***
Retention test * Non-native	1.91 (0.77)*	2.22 (1.00)*
Years behind	-1.69 (0.47)***	-2.79 (0.46)***
Fluency	0.29 (0.02)***	0.25 (0.02)***
Posttest * Fluency	-0.12 (0.02)***	-0.10 (0.02)***
Retention test * Fluency	-0.14 (0.02)***	-0.12 (0.02)***
STRAT+SA (pretest)	0.50 (1.12)	2.27 (0.94)*
STRAT+CA (pretest)	1.24 (1.10)	-0.01 (0.87)
STRAT (pretest)	1.38 (0.84)	
Posttest * STRAT+SA	0.58 (0.78)	0.69 (1.09)
Posttest * STRAT+CA	0.62 (0.74)	2.13 (1.00)*
Posttest * STRAT	0.53 (0.58)	
Retention test * STRAT+SA	1.16 (0.87)	0.07 (1.30)
Retention test * STRAT+CA	-1.11 (0.83)	1.15 (1.19)
Retention test * STRAT	-0.70 (0.65)	
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	1.18 (0.58)*	1.27 (0.82)
$\sigma_{v0vpost}$		-1.55 (0.87)
$\sigma^2_{vpost}$		2.12 (1.09)
$\sigma_{v0vret}$		-1.90 (1.02)
$\sigma_{vpostvret}$		2.63 (1.21)*
$\sigma^2_{vret}$		3.58 (1.57)*
Level 2		
$\sigma^2_{\mu0}$	24.33 (1.76)***	25.37 (1.88)***
$\sigma_{\mu0\mu post}$	-14.29 (1.39)***	-16.66 (1.59)***
$\sigma^2_{\mu post}$	22.00 (1.60)***	24.41 (1.86)***
$\sigma_{\mu0\mu ret}$	-15.87 (1.56)***	-15.92 (1.57)***
$\sigma_{\mu post\mu ret}$	13.93 (1.46)***	15.43 (1.55)***
$\sigma^2_{\mu ret}$	26.27 (1.98)***	23.18 (1.84)***
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

As concerns the effects of the experimental conditions, Table 7.1 reveals no significant differential evolution of second graders in the first study's experimental STRAT, STRAT+SA, or STRAT+CA conditions in comparison with the control group students' evolution. These results are rather different from the previous results by

means of the two-level regression analyses on students' post- and retention test data reported in Chapter 2. In these latter analyses, significant posttest effects were found for the STRAT and STRAT+CA condition, entailing that by the end of the school year second-grade students in these conditions had made significantly more progress in reading comprehension than children in the control group. Notwithstanding the nonsignificant results revealed by the three-level repeated measures analysis on the first study's data, in the second study the pre- to posttest change of second graders in the STRAT+CA condition appeared to be significantly better than that of the control group students. More specifically, the estimate corresponds with an effect size of 0.42 standard deviation. With regard to the progress over the whole period from pretest to retention test, the difference between the STRAT+CA and control condition was no longer significant. Pairwise comparisons of the experimental conditions in both studies revealed no significant differences at neither the post- nor the retention test.

We can conclude that with respect to short-term effects the second study seems to yield better results than the first study, at least with regard to the STRAT+CA condition. This does not contradict our expectations since by the time of the second study teachers got more experienced with the experimental treatment. The fact that this improvement was particularly found for the STRAT+CA condition points out the higher potential of this condition. The finding that the second and the first study did not differ with respect to the absence of long-term effects confirms the idea that for these young students long-term effects are only likely to appear if the experimental treatment is continued.

As concerns the effects of the fifth-grade experimental conditions, Table 7.2 reports that the progress of students in the first study's STRAT+CA and STRAT conditions was significantly higher in comparison with the control group students' advancement. Effect sizes are 0.36 and 0.31 standard deviation respectively. Moreover, both conditions appeared to keep growing, resulting in a significantly larger progress from pretest to retention test with effect sizes of 0.75 and 0.46 respectively. In addition, a pairwise comparison of the three experimental conditions also indicated a significantly larger progress from pre- to retention test of students in the STRAT+CA condition in comparison with children in the STRAT+SA condition, with an effect size of 0.56 standard deviation. With regard to the findings of the second study, Table 7.2 reveals that the progress from pre- to posttest in both the STRAT+SA and STRAT+CA condition was significantly higher than in the control condition. Effect sizes are 0.21 and 0.28 standard deviation respectively. With regard to the progress over the whole period from pretest to retention test only students in the STRAT+CA condition experienced significant differential learning gains with an effect size of 0.42 standard deviation. The differential slope of the STRAT+SA condition, with an effect size of 0.28 standard deviation, only reached a marginal significance level. A pairwise comparison of the experimental conditions showed no significant difference between the STRAT+SA and STRAT+CA condition's progress from pretest to post- or retention test.

Table 7.2

*Model Estimates for the Repeated Measures Three-Level Analyses of the Fifth-Grade Reading Comprehension Achievement Scores for Both Studies*

Parameter	Study	
	1 <sup>st</sup>	2 <sup>nd</sup>
<i>Fixed</i>		
Intercept	39.79 (2.02)***	42.21 (2.01)***
Posttest (control group)	0.92 (1.66)	0.94 (1.01)
Retention test (control group)	3.33 (1.92)	3.03 (1.97)
Boy		-2.64 (1.28)*
Non-native	-11.52 (2.48)***	-7.50 (2.71)**
Years behind		-9.71 (1.36)***
STRAT+SA (pretest)	1.99 (3.03)	1.10 (2.42)
STRAT+CA (pretest)	-8.05 (3.18)*	-1.88 (2.50)
STRAT (pretest)	-0.52 (2.60)	
Posttest * STRAT+SA	3.96 (2.54)	3.02 (1.28)*
Posttest * STRAT+CA	5.60 (2.64)*	4.35 (1.32)***
Posttest * STRAT	4.84 (2.17)*	
Retention test * STRAT+SA	2.86 (2.95)	4.55 (2.59)
Retention test * STRAT+CA	11.62 (3.05)***	6.59 (2.65)*
Retention test * STRAT	7.11 (2.52)**	
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	13.01 (6.81)	8.19 (5.21)
$\sigma_{v0vpost}$	-3.95 (4.26)	
$\sigma^2_{vpost}$	11.30 (4.78)*	
$\sigma_{v0vret}$	10.86 (5.00)*	-0.94 (4.08)
$\sigma_{vpostvret}$	-4.91 (3.95)	
$\sigma^2_{vret}$	14.63 (6.44)*	14.13 (6.22)*
Level 2		
$\sigma^2_{\mu0}$	189.41 (13.04)***	220.41 (15.36)***
$\sigma_{\mu0\mu post}$	-46.34 (6.68)***	-70.72 (8.27)***
$\sigma^2_{\mu post}$	83.93 (5.96)***	100.58 (7.19)***
$\sigma_{\mu0\mu ret}$	-52.53 (8.07)***	-74.68 (10.01)***
$\sigma_{\mu post\mu ret}$	50.18 (5.74)***	67.59 (7.19)***
$\sigma^2_{\mu ret}$	119.67 (8.70)***	143.65 (10.84)***
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

Generally, the results of the second study seem to confirm the results of the first study, especially with respect to the positive short- and long-term effects of the STRAT+CA condition. However, the effect sizes seem to be smaller in the latter study. Presumably, this finding can be explained to a large extent by the low average pretest score for the STRAT+CA condition in the first study. This low average pretest score was quite unexpected, since all efforts were done to match teachers and classes in the different conditions, based on fourth-grade data from the school year preceding the first experimental treatment. The results of the first study seem to suggest that the students in the cross-age peer tutoring condition were able to catch up during fifth grade and together with the positive effect of the experimental treatment this yielded a quite high effect size. Probably the effect sizes found in the second study are the more realistic ones. With regard to the STRAT+SA condition the second study shows better results

than the first study, since a positive significant short-term effect was found, whereas previously no significant effect was revealed for this condition. No comparison can be made for the strategies-only condition of course since this condition was not included in the second study.

### *Effects on Students' Reports of the Use of Reading Strategies*

With regard to second and fifth graders' reports of strategy use, similar models were built up. Only for second graders significant intervention effects were found. Table 7.3 presents the final repeated measures models for second-grade results in both studies. The second study's model is identical to the model reported in Chapter 6. The analyses for the first study by means of two-level regression analyses revealed no significant effects of the STRAT, STRAT+SA, and STRAT+CA condition, neither at posttest (resp.  $\chi^2 = 0.44$ ,  $df = 1$ ,  $p = .505$ ;  $\chi^2 = 3.19$ ,  $df = 1$ ,  $p = .074$ ;  $\chi^2 = 1.03$ ,  $df = 1$ ,  $p = .311$ ), nor at retention test (resp.  $\chi^2 = 0.13$ ,  $df = 1$ ,  $p = .717$ ;  $\chi^2 = 1.72$ ,  $df = 1$ ,  $p = .190$ ;  $\chi^2 = 0.17$ ,  $df = 1$ ,  $p = .677$ ). However, as Table 7.3 points out, the revised analysis of the data by means of a three-level repeated measures model actually did reveal significant results. In fact, these first study's results were quite similar to the results of the second study.

Table 7.3 indicates that, after controlling for students' gender, neither for the first, nor for the second study significant pretest differences between the different research conditions were found. Compared to the beginning of the school year control group students seem to report less frequent use of reading strategies at the time of the post- and retention test. But, contrary to the control group students all students in the experimental conditions reported more frequent use of reading strategies at the end of second grade. More specifically, in the first study effect sizes of respectively 0.39, 0.45, and 0.33 standard deviation were found for the STRAT+SA, STRAT+CA, and STRAT condition. In the second study effect sizes were respectively 0.42 standard deviation for the STRAT+SA condition and 0.71 standard deviation for the STRAT+CA condition. Only for the second study's STRAT+CA condition the difference with the control group students remained significant at the retention test, with an effect size of 0.62 standard deviation. Moreover, a pairwise comparison of the experimental conditions in the second study revealed a significant difference between the STRAT+SA and STRAT+CA condition as well. More specifically, a significantly larger increase in the use of reading strategies was reported by second graders in the STRAT+CA condition, both at post- and retention test. Effect sizes are respectively 0.29 and 0.31 standard deviation.

In general, the results for the second study appear to confirm the results from the first study. However, the higher effect sizes in the latter study and the appearance of a long-term effect for the STRAT+CA condition show an improvement of the experimental treatment effects. More particularly, especially the effects of the STRAT+CA condition appear to have improved from the first to the second study.



Table 7.3

*Model Estimates for the Repeated Measures Three-Level Analyses of the Second-Grade Measures of the Use of Reading Comprehension Strategies for Both Studies*

Parameter	Study	
	1 <sup>st</sup>	2 <sup>nd</sup>
<i>Fixed</i>		
Intercept	2.86 (0.07)***	2.83 (0.07)***
Posttest (control group)	-0.15 (0.05)**	-0.16 (0.05)**
Retention test (control group)	-0.10 (0.05)	-0.10 (0.05)*
Boy	-0.14 (0.04)***	-0.09 (0.03)**
STRAT+SA (pretest)	-0.11 (0.10)	-0.11 (0.09)
STRAT+CA (pretest)	-0.06 (0.12)	-0.14 (0.09)
STRAT (pretest)	-0.14 (0.09)	
Posttest * STRAT+SA	0.19 (0.08)*	0.19 (0.07)**
Posttest * STRAT+CA	0.24 (0.09)**	0.32 (0.06)***
Posttest * STRAT	0.16 (0.07)*	
Retention test * STRAT+SA	-0.06 (0.08)	0.14 (0.07)
Retention test * STRAT+CA	0.03 (0.09)	0.27 (0.07)***
Retention test * STRAT	0.12 (0.07)	
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	0.02 (0.01)*	0.02 (0.01)*
Level 2		
$\sigma^2_{\mu0}$	0.19 (0.01)***	0.15 (0.02)***
$\sigma_{\mu0\mu\text{post}}$	-0.12 (0.01)***	-0.13 (0.02)***
$\sigma^2_{\mu\text{post}}$	0.33 (0.02)***	0.27 (0.02)***
$\sigma_{\mu0\mu\text{ret}}$	-0.15 (0.02)***	-0.14 (0.02)***
$\sigma_{\mu\text{post}\mu\text{ret}}$	0.19 (0.02)***	0.17 (0.02)***
$\sigma^2_{\mu\text{ret}}$	0.33 (0.02)***	0.29 (0.02)***
$\sigma_{\mu0\mu\text{boy}}$		0.03 (0.01)*
$\sigma_{\mu\text{post}\mu\text{boy}}$		-0.01 (0.02)
$\sigma_{\mu\text{ret}\mu\text{boy}}$		-0.02 (0.02)
$\sigma^2_{\mu\text{boy}}$		0.00 (0.00)
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

### *Effects on Self-Efficacy Related Thoughts*

As regards students' thoughts about reading proficiency and success or failure attributions, statistically significant effects for second graders were found for thoughts relating to success attributions and positive self-efficacy perceptions towards reading. For fifth graders, on the other hand, significant effects were found for thoughts relating to failure attributions and negative self-efficacy perceptions. The results of the analyses on the second- and fifth-grade data for both studies are respectively presented in Table 7.4 and 7.5. For the second study both models are identical to the models reported in Chapter 6. For students' self-efficacy related thoughts in the first study, revised models were estimated by means of three-level repeated measures analyses.

With regard to second graders' estimates, Table 7.4 reports significant pretest differences in the second study after taking into account students' gender. More specifically, it was found that second graders in both the STRAT+SA and STRAT+CA

condition reported significantly less positive self-efficacy related thoughts than their control group peers. No pretest differences were found in the first study. In both studies the incidence of positive self-efficacy related thoughts as reported by the control group students themselves decreased significantly from pretest to posttest and remained more or less stable from posttest to retention test.

As concerns second-grade experimental conditions' differential change from pre- to post- and retention test in comparison with the control group's evolution, Table 7.4 reveals that no significant effects were found by means of a three-level repeated measures model for the first study's data. This result confirms the findings of the two-level regression analyses on the same data, where no significant differences were found between the STRAT, STRAT+SA, and STRAT+CA condition in comparison with the control group at post- (resp.  $\chi^2 = 0.60$ ,  $df = 1$ ,  $p = .439$ ;  $\chi^2 = 0.45$ ,  $df = 1$ ,  $p = .502$ ;  $\chi^2 = 0.55$ ,  $df = 1$ ,  $p = .458$ ) or retention test (resp.  $\chi^2 = 0.51$ ,  $df = 1$ ,  $p = .475$ ;  $\chi^2 = 0.13$ ,  $df = 1$ ,  $p = .718$ ;  $\chi^2 = 1.53$ ,  $df = 1$ ,  $p = .216$ ). For the second study, however, it was found that both the STRAT+SA and STRAT+CA conditions' change from pre- to posttest was significantly higher than the control group's evolution. Effect sizes are 0.44 and 0.56 standard deviation respectively. Contrary to the control group students reporting significantly less thoughts with regard to positive self-efficacy related thoughts at posttest, the experimental students' reports remained more or less the same. With regard to the progress over the whole period from pretest to retention test, the significant difference with respect to the control group disappeared for both conditions. Neither at posttest, nor at retention test pairwise comparisons of the experimental conditions revealed significant differences between the STRAT+SA and STRAT+CA condition.

Generally, it can be concluded that for the second-grade students the results from the second study seem to be better than the results from the first study. However, part of the effects that were found in the second study could be attributed to the lower pretest scores that were found for the experimental conditions.

With regard to the fifth-grade models, Table 7.5 indicates that, after controlling for students' mother tongue, neither for the first, nor for the second study significant pretest differences between the different research conditions were found. The incidence of negative self-efficacy related thoughts as reported by the control group students remained stable from the start to the end of the school year and till six months after finishing the treatment.

As concerns the experimental conditions' effects, Table 7.5 reveals a significantly larger decrease from pretest to retention test in the preoccupation with negative self-efficacy related thoughts as reported by students in the first study's STRAT+SA condition in comparison with students in the control group. More specifically an effect size of 0.36 standard deviation was found. This revised analysis by means of a three-level repeated measures model yields rather different results than the initial analyses reported in Chapter 2. In these latter analyses it was revealed that by the end of the school year children in the STRAT+CA condition were significantly less occupied with negative self-efficacy related thoughts compared to students in the control and STRAT+SA condition. Moreover, the significant difference with the control group

endured at retention test. In addition, at retention test STRAT+SA students reported significantly less negative thoughts than the control group students as well.

From the three-level repeated measures model on the second study's data, a significant STRAT+SA effect arose at posttest, with an effect size of 0.40 standard deviation. Moreover, the differential pre- to posttest evolution of the STRAT+SA condition was also significantly different from the STRAT+CA condition's evolution, with an effect size of 0.29 standard deviation. At retention test, however, the significant difference with the control condition became only marginally significant and the difference between the STRAT+SA and STRAT+CA condition disappeared.

For the effects discussed here, the results from the second study can not be easily compared with those from the first study. The common finding is that "positive" effects were only found for the STRAT+SA condition, but whereas the first study only shows a long-term effect without any short-term effect preceding it, the second study shows mainly a short-term effect, followed by a marginally significant long-term effect.

Table 7.4

*Model Estimates for the Repeated Measures Three-Level Analyses of the Second-Grade Measures of Positive Self-Efficacy Related Thoughts for Both Studies*

Parameter	Study	
	1 <sup>st</sup>	2 <sup>nd</sup>
<i>Fixed</i>		
Intercept	3.37 (0.09)***	3.45 (0.09)***
Posttest (control group)	-0.33 (0.09)***	-0.33 (0.09)***
Retention test (control group)	-0.32 (0.13)*	-0.31 (0.09)***
Boy		-0.13 (0.05)*
STRAT+SA (pretest)	-0.11 (0.14)	-0.27 (0.13)*
STRAT+CA (pretest)	0.06 (0.16)	-0.34 (0.12)**
STRAT (pretest)	-0.11 (0.12)	
Posttest * STRAT+SA	0.18 (0.14)	0.31 (0.13)*
Posttest * STRAT+CA	-0.18 (0.15)	0.39 (0.12)***
Posttest * STRAT	0.00 (0.12)	
Retention test * STRAT+SA	0.01 (0.19)	0.11 (0.12)
Retention test * STRAT+CA	-0.31 (0.22)	0.12 (0.11)
Retention test * STRAT	-0.05 (0.17)	
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	0.03 (0.01)*	0.03 (0.01)*
$\sigma_{v0vret}$	-0.01 (0.01)	
$\sigma^2_{vret}$	0.05 (0.03)*	
Level 2		
$\sigma^2_{\mu0}$	0.40 (0.03)***	0.43 (0.03)***
$\sigma_{\mu0\mu post}$	-0.29 (0.03)***	-0.36 (0.04)***
$\sigma^2_{\mu post}$	0.91 (0.07)***	0.86 (0.07)***
$\sigma_{\mu0\mu ret}$	-0.33 (0.03)***	-0.35 (0.04)***
$\sigma_{\mu post\mu ret}$	0.50 (0.05)***	0.49 (0.05)***
$\sigma^2_{\mu ret}$	0.80 (0.06)***	0.76 (0.06)***
Level 1		
$\sigma^2_{\epsilon0}$	0.00 (0.00)	0.00 (0.00)

*Note.* Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

Table 7.5

*Model Estimates for the Repeated Measures Three-Level Analyses of the Fifth-Grade Measures of Negative Self-Efficacy Related Thoughts for Both Studies*

Parameter	Study	
	1 <sup>st</sup>	2 <sup>nd</sup>
<i>Fixed</i>		
Intercept	1.88 (0.05)***	1.88 (0.05)***
Posttest (control group)	0.04 (0.05)	0.04 (0.05)
Retention test (control group)	-0.05 (0.05)	-0.05 (0.05)
Non-native	0.23 (0.09)**	-0.25 (0.09)**
STRAT+SA (pretest)	0.07 (0.07)	0.10 (0.06)
STRAT+CA (pretest)	0.02 (0.08)	0.03 (0.07)
STRAT (pretest)	0.08 (0.07)	
Posttest * STRAT+SA	-0.05 (0.07)	-0.22 (0.06)***
Posttest * STRAT+CA	-0.09 (0.08)	-0.06 (0.06)
Posttest * STRAT	-0.09 (0.06)	
Retention test * STRAT+SA	-0.18 (0.07)*	-0.12 (0.06)
Retention test * STRAT+CA	-0.10 (0.08)	-0.11 (0.06)
Retention test * STRAT	-0.11 (0.06)	
<i>Random</i>		
Level 3		
$\sigma^2_{v0}$	0.00 (0.00)	0.00 (0.00)
Level 2		
$\sigma^2_{\mu0}$	0.28 (0.02)***	0.26 (0.02)***
$\sigma_{\mu0\mu\text{post}}$	-0.09 (0.01)***	-0.11 (0.01)***
$\sigma^2_{\mu\text{post}}$	0.20 (0.01)***	0.25 (0.02)***
$\sigma_{\mu0\mu\text{ret}}$	-0.12 (0.01)***	-0.11 (0.01)***
$\sigma_{\mu\text{post}\mu\text{ret}}$	0.12 (0.01)***	0.14 (0.01)***
$\sigma^2_{\mu\text{ret}}$	0.22 (0.02)***	0.24 (0.02)***
Level 1		
$\sigma^2_{\varepsilon0}$	0.00 (0.00)	0.00 (0.00)

Note. Per cell: regression coefficient and standard error

\* $p < .05$  \*\* $p < .01$  \*\*\* $p < .001$

### *Effects on Reading Fluency Achievement, Sociometric Rating Score, Self-Concept, and Attitude towards Reading*

As was reported in Chapter 4, the impact of the first study's experimental interventions on second graders' reading fluency development was immediately analyzed by means of a three-level repeated measures model. The results more specifically revealed that the STRAT+CA condition's pre- to posttest growth in reading fluency was significantly higher than the progress of the control condition. The effect size reached a magnitude of 0.37 standard deviation. Contrary to the positive results of the first study, no significant effects of the experimental interventions were found for the second study.

By means of a sociometric rating procedure (Singleton & Asher, 1977) second and fifth graders evaluated how much they like to collaborate with each of their classmates on a four-point Likert-type response format. The individual rating score for each student is computed by averaging out the ratings obtained from all classmates. In neither studies significant intervention effects were found on second or fifth graders' obtained rating scores.

To measure fifth graders' self-concept, the "Competentiebelevingsschaal voor Kinderen (CBSK)" (Veerman, Straathof, Treffers, Van den Bergh, & ten Brink, 1997), a Dutch version of the Self-Perception Profile for Children (Harter, 1985), was administered at pre- and posttest in both studies. More specifically, the scales measure children's self-perceptions in relation to specific domains of one's life (scholastic competence, athletic competence, social acceptance, physical appearance, and behavioral conduct) and a separate facet of global self-worth. The CBSK is a 36-item self-report questionnaire divided into the six subscales mentioned above. For example, to assess self-perceived scholastic competence, children were asked to first choose which of the following sentences sounded more like them: "Some children feel that they are very good at their school work" or "Other children worry about whether they can do the school work assigned to them." After the children had decided which sentence sounded more like themselves, they were then asked to judge whether the statement was "really true" or "sort of true" for them. All questions followed the same format. Because the questionnaire was not suitable for second graders, this instrument was not used with this age group. In neither studies significant intervention effects were found on the different scales of the questionnaire.

As concerns the assessment of students' attitude towards reading, children completed a Dutch Reading Attitude Scale with 27 items (Aarnoutse, 1996) at pre- and posttest in both studies. Fifth graders read and completed the questionnaire individually. In second grade all items were read out loud by the test administrator and completed individually by the students. Reading attitude can be unfolded in three components, that is children's knowledge and experience with reading, their positive or negative appreciation of reading and reading material, and their tendency to read. The Reading Attitude Scale refers to these different elements, but focuses chiefly on the affective aspect. The score on the scale is measured as the number of positive judgments on the 27 questions. In neither studies significant intervention effects were found on second or fifth graders' attitude towards reading.

### *Conclusion*

Generally, it can be noticed that significant intervention effects were found with regard to second and fifth graders' reading comprehension achievement and self-efficacy related thoughts towards reading, as well as with regard to second graders' reading fluency skills and reported strategy use. No significant findings were discerned in students' sociometric rating score, self-concept, and attitude towards reading. To structuralize the recapitulation of the findings, Table 7.6 presents an overview of the significant effects and corresponding effect sizes revealed in the studies.

For most of the dependent variables it can be concluded that the results of the second study generally confirm those of the first study. This is particularly true for the effect of both peer tutoring conditions on the use of reading strategies in second-grade students and the effect of the cross-age peer tutoring condition on reading comprehension in fifth grade. It is also true for the absence of any treatment effect on

sociometric rating scores, children's self concept profile, and attitude towards reading as measured by student rating scales.

Table 7.6

*Overview of the Significant Effects and Corresponding Effect Sizes Revealed in Both Studies*

	Study 1 <sup>a</sup>		Study 2 <sup>a</sup>	
	Posttest	Retention test	Posttest	Retention test
<i>Second grade</i>				
Reading fluency	STRAT+CA>C (ES=0.37)			
Reading comprehension			STRAT+CA>C (ES=0.42)	
Use of reading strategies	STRAT>C (ES=0.33)			
	STRAT+SA>C (ES=0.39)		STRAT+SA>C (ES=0.42)	
	STRAT+CA>C (ES=0.45)		STRAT+CA>C (ES=0.71)	STRAT+CA>C (ES=0.62)
			STRAT+CA > STRAT+SA (ES=0.29)	STRAT+CA > STRAT+SA (ES=0.31)
Positive self-efficacy related thoughts			STRAT+SA>C (ES=0.44)	
			STRAT+CA>C (ES=0.56)	
<i>Fifth grade</i>				
Reading comprehension	STRAT>C (ES=0.31)	STRAT>C (ES=0.46)		
			STRAT+SA>C (ES=0.21)	
	STRAT+CA>C (ES=0.36)	STRAT+CA>C (ES=0.75)	STRAT+CA>C (ES=0.28)	STRAT+CA>C (ES=0.42)
		STRAT+CA > STRAT+SA (ES=0.56)		
Negative self-efficacy related thoughts		STRAT+SA<C (ES=0.36)	STRAT+SA<C (ES=0.40)	
			STRAT+SA < STRAT+CA (ES=0.29)	

<sup>a</sup> STRAT = strategies-only condition; STRAT+SA = same-age peer tutoring condition; STRAT+CA = cross-age peer tutoring condition; C = control condition

With regard to dependent variables for which significant effects of the experimental conditions were found, in many cases the effects appeared or were more pronounced in

the second study. This is more particularly true in second grade for the effect of both peer tutoring conditions on the use of reading strategies and positive self-efficacy related thoughts and for the effect of the cross-age peer tutoring condition on reading comprehension. It is also true for the positive short-term effect of the same-age peer tutoring condition on reading comprehension in fifth grade. That results in the second study generally appear to be better could be expected since most teachers that were involved in the second study also had been involved in the first study. By the time the experimental treatment was done for the second time, they had one year experience in at least one of the cornerstones of the innovation, namely the explicit teaching of reading strategies.

## **General Discussion**

The present section deals with a general discussion of the two successive studies' results, in which attention is mainly paid to answering the research questions as formulated in Chapter 1. As concerns the effects on student outcomes, the review is restricted to the general discussion of the results obtained by the hierarchical repeated measures analyses reported in the previous section.

### *Discussion with regard to the First Research Question*

With regard to the first research question, it can be inferred from both studies' results that the experimental treatments in general influenced elementary school students' reading comprehension and fluency achievement, reports of reading strategy use, and self-efficacy related thoughts. Below, we go into a more in-depth discussion of the results.

### *Reading Comprehension*

As concerns students' reading comprehension, the findings generally accord with previous research, emphasizing the educational benefits of explicit reading strategies instruction (e.g., Alfassi, 1998; Baumann, Seifert-Kessel, & Jones, 1992; Block, 1993; Brand-Gruwel, Aarnoutse, & Van Den Bos, 1998; Brown, Pressley, Van Meter, & Schuder, 1996; Dole, Brown, & Trathen, 1996; Duffy et al., 1987; Fuchs, Fuchs, Mathes, & Simmons, 1997; Haller, Child, & Walberg, 1988; Kelly & Moore, 1994; Klingner & Vaughn, 1996; Klingner, Vaughn, & Schumm, 1998; Palincsar & Brown, 1984; Pressley, 2000; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989) and peer-led interaction about texts (e.g., Alfassi, 1998; Almasi, 1996; Almasi, McKeown, & Beck, 1996; Brown et al., 1996; Fuchs & Fuchs, 2000; Fuchs et al., 1997; Johnson-Glenberg, 2000; Kelly & Moore, 1994; Klingner & Vaughn, 1996; Klingner et al., 1998; Mathes, & Fuchs, 1994; Palincsar & Brown, 1984; Pressley et al., 1992; Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995) in promoting reading comprehension achievement. Contrary to other studies (e.g., Alfassi, 1998; De Corte, Verschaffel, & Van De Ven, 2001; Walraven, 1995), significant results were found on standardized reading comprehension tests instead of on experimenter-designed tests, which even strengthen and underlines the value of the experimental treatments.

For second graders, the results chiefly document the feasibility of fostering reading comprehension achievement by supplying students with explicit instruction in reading strategies, followed by practice in cross-age peer tutoring activities led by a fifth-grade tutor. No similar effects could, however, be detected for students practicing the explicitly taught reading strategies in either teacher-led whole-class activities or second-grade reciprocal same-age dyads. Both these conditions bear some resemblance to Palincsar and Browns' (1984) reciprocal teaching approach, in which students acquire reading strategies during teacher-guided practice in small groups taking the format of a dialogue, where the teacher and the students take turns leading the discussion. In this respect, the absence of significant extra learning gains for the strategies-only and same-age tutoring conditions parallels the findings of a review of 16 reciprocal teaching studies (Rosenshine & Meister, 1994), which revealed significant effects of the reciprocal teaching approach from grades four to adult education, but nonsignificant effects for younger students. The nonsignificant result of second-grade reciprocal same-age tutoring, however, contrasts with other research studying the efficacy of a reciprocal same-age peer tutoring program in reading on a variety of age groups, including second graders (Fuchs et al., 1997; Simmons et al., 1995). Nevertheless, it has to be noted in this respect that in those studies no separate analyses were done on second graders' data.

As to the cross-age tutoring condition, it can be hypothesized that the significant findings for second grade can be attributed to two constituent elements of the experimental treatment, namely the individualized practice and assistance provided to the second graders, and the fact that fifth graders appear to be sufficiently competent in offering support to practice strategic reading effectively. In the other two nonsignificant experimental conditions, these two essential conditions are not fulfilled simultaneously. The strategies-only condition is characterized by proficient support by the teacher, but no individualized attention is offered. On the other hand, the reciprocal same-age peer tutoring condition supplies individualized assistance, but it can be assumed that second graders are not able yet to supervise and tutor the practice of reading strategies effectively.

For fifth graders, the combination of both studies' reading comprehension results confirm that explicit reading strategies instruction, supplemented by either teacher- or peer-mediated practice, foster comprehension achievement more than traditional reading comprehension instruction. In this respect, it is important to notice that the effect sizes of the reported results, in particular of the strategies-only and cross-age conditions, range from 0.28 to 0.75 standard deviation. These compare favorably with the treatment effects on standardized tests reported in Rosenshine and Meister's (1994) review of reciprocal teaching, one of the most widespread and prominent strategy-instruction programs. Moreover, except for students practicing the reading strategies by alternating between the role of tutor and tutee in reciprocal same-age tutoring activities, the effectiveness of the experimental treatments is especially corroborated by the significant long-term effects that were found until six months after the end of the experimental interventions, with effect sizes between 0.42 and 0.75 standard deviation. Taken into account the magnitude of the posttest and retention test effects, it can be concluded that the fifth-grade tutors seem to profit most from being engaged in



the cross-age peer tutoring condition. This finding is especially noticeable since in cross-age tutoring activities fifth-grade tutors are working with texts on their tutees' level, that is second-grade level instead of fifth-grade level. Nevertheless, their learning gains outperform other fifth graders' growth and they effectuate the most persistent long-term progress in reading comprehension.

A possible explanation of fifth graders' considerable improvement in reading comprehension, while in fact they only have been monitoring and regulating second graders' reading of second-grade level texts, relates to the nature of tutoring a peer in connection to the opportunity to practice and acquire metacognitive monitoring and regulation skills. The tutoring task requires tutors to focus their attention on the tutee's reading, to superintend their reading process attentively, to monitor their understanding, and to endeavor to foster comprehension by having their tutees apply relevant, well-chosen reading strategies. The fact not having to read themselves makes it easier for tutors to give full attention to these metacognitive skills. In other words, it may be easier and more appropriate to practice the metacognitive skills involved in the process of reading comprehension by monitoring and regulating someone else's reading process than one's own reading process, more particularly if this someone else is a younger student with a lower reading level. In this respect, the opportunity to practice the explicitly taught reading strategies while functioning as a tutor for a younger student can be seen as a powerful learning environment for enhancing the independent and appropriate application of reading strategies and for acquiring skills in metacognitive monitoring and regulation of one's own reading process, leading to a long-term improvement of reading comprehension achievement.

From a theoretical point of view, it can be asserted that the positive results of the peer tutoring conditions on second- and fifth-grade students' reading comprehension achievement are consistent with Vygotsky's perspective on socially mediated learning, which emphasizes the importance of communication between learners and argues that the process of internalization begins in a social context and then develops within the individual (Vygotsky, 1978). More particularly, in analogy with Palincsar and Brown (1984) and Brown et al. (1996) it can be argued that participation in the reading dyads' interaction and discussions that involve reading strategy use leads to internalization and consistently adaptive use of strategic processing by the team members.

### *Reading Fluency*

Notwithstanding the fact that the essence of the innovative instructional approach focused on reading comprehension, it was hypothesized that the considerable amount of reading practice shared with peers in regular peer tutoring sessions, would also ameliorate accurate reading with appropriate rate. Therefore, the side effect of the interventions on students' reading fluency achievement was also explored. In this respect, data were solely collected in the second-grade classes, since it is recognized that reading fluency is generally well developed at the end of the third grade (Bast & Reitsma, 1998; Sticht & James, 1984). The findings of the first study reveal that compared to conventional instruction in the control condition, students practicing strategic reading in regularly organized cross-age tutoring sessions led by a fifth grader, grow significantly more on reading fluency in the course of second grade. In

addition, the effects of reading in reciprocal same-age dyads display a similar, though not statistically significant trend. In accordance with the results of previous studies (Simmons et al., 1995; Walraven, 1995), the absence of an analogous result for students in the strategies-only condition demonstrates that receiving explicit reading strategies instruction without extra opportunities to practice oral reading in peer tutoring dyads does not result in increased reading fluency skills. Generally, the results corroborate the findings of previous research confirming the positive impact of readings shared with peers on oral fluent reading (Fuchs et al., 1997; Simmons, Fuchs, Fuchs, Hodge, & Mathes, 1994; Simmons et al., 1995). Contrary to those studies documenting the efficacy of same-age peer tutoring, the present results however only reveal a significant effect for reading peer tutoring activities led by an older, fifth-grade tutor. This confirms the result of a study of Dixon-Kraus (1995), which indicated that reading with a peer of comparable ability is not likely to improve fluency as much as reading with a more able reader. The fact that significant effects were only found for second graders tutored by an older student, leads us to suspect that the assistance and corrective feedback provided by fifth graders is more appropriate than the assistance and feedback given by second-grade peers.

Unfortunately, the second study without significant treatment effects does not confirm the positive findings of study one. From the overview in Table 7.6 it seems as if the significant effect on students' reading fluency in the first study is replaced by the significant reading comprehension achievement result in the second study. However, it is not clear that these two effects are really related or what a possible explanation for the shift could be.

### *Reports of Strategy Use*

With regard to students' reported use of reading strategies, the findings of both studies only revealed significant effects of the treatments for second graders. More specifically, it was shown that explicit reading strategies instruction followed by regular teacher- or peer-mediated practice enhanced second graders' strategy use as reported by themselves significantly more than traditional reading comprehension instruction. These second-grade results are congruous with previous studies, which applied reading strategies tests, questionnaires or interviews (e.g., Brand-Gruwel et al., 1998; Brown et al., 1996; De Corte et al., 2001; Duffy et al., 1987; Klingner & Vaughn, 1996; Klingner et al., 1998; Palincsar & Brown, 1984; Pressley et al., 1989; Walraven, 1995), and suggest that by receiving explicit strategies instruction and practice, second graders become more aware of the benefits of using reading strategies during the reading process to enhance the likeliness of comprehension. When comparing the magnitude and the maintenance of the effects, fifth graders functioning as tutor for second graders in cross-age dyads seem most effective in accomplishing this awareness, even in the long term.

Contrary to the expectations on the basis of the aforementioned research and the reported positive findings for all experimental interventions on fifth graders reading comprehension skills, no significant results were recorded with regard to fifth graders' reports of reading strategy use. It can be hypothesized that this failure to establish significant progress in fifth graders' strategic reading reports, can be attributed to the fact that the applied questionnaire was not an appropriate measurement instrument for

this age group. More specifically, it is assumed that the questionnaire statements with respect to the use of strategies before, while, and after reading speak for themselves, causing a tendency to social desirability in answering the questionnaire. In this respect, reading strategies tests, thinking-aloud protocols, or stimulated recall techniques will be interesting supplements to the applied questionnaire to shed light on students' reading behavior, application of reading strategies, and metacognitive regulation of the reading process (Veenman & van Hout-Wolters, 2002). Moreover, this triangulation of collecting data would give in to the objection that children's reportage of strategy use is not necessarily consistent with their actual behavior while reading.

### *Reports of Self-Efficacy Related Thoughts*

As concerns students' thoughts about their reading proficiency, no concurrent results were found for second and fifth graders. For second graders, the results document that control group students report significantly less thoughts with regard to positive self-efficacy related thoughts by the end of second grade in comparison with the start of the school year. On the contrary, second graders' practicing strategic reading in peer tutoring dyads, report more or less the same amount of positive thoughts. Notwithstanding these significant effects, one has to take into account that the experimental conditions' pretest scores were not equivalent to the control group's scores, which imports a serious threat to the internal validity of these results. In addition, it is also important to mention that a decrease in reported positive self-efficacy related thoughts, as could be deduced from control group students' reports, does not necessarily mean that students' self-esteem grow worse. Strictly spoken it merely indicates being less preoccupied with thoughts about that.

As regards fifth graders' preoccupation with self-efficacy related thoughts, the findings only reveal effects for negative self-efficacy related thoughts. The combination of the results of both studies confirm that by the end of the school year and at retention test the reports of students practicing strategic reading in fifth-grade reciprocal same-age dyads display a significant decrease in negative self-efficacy related thoughts. Since it can be assumed that high scores for negative self-efficacy related thoughts indicate the presence of low self-esteem with regard to reading ability, these results indicate that opportunities to alternate between the roles of tutor and tutee in same-age dyads have a positive impact on fifth graders' self-esteem and perceived competence. More specifically, students become more confident about their own reading competence and are less preoccupied with doubts.

Unlike the results of the same-age tutoring condition, the effects of the cross-age peer tutoring condition are not univocal. By the end of the school year the decline in the reports of negative self-efficacy related thoughts is significantly smaller than that of their peers in the same-age condition. However, six months after the end of the experimental treatment, the differential change from pretest to retention test with respect to the control group becomes marginally significant and the significant difference as compared to the same-age condition disappears. Clearly, additional research is necessary to shed light on these ambiguous results. Again, stimulated recall techniques or thinking-aloud protocols in response to exposure to reading comprehension tasks with divergent levels of difficulty could be an interesting

supplement to assess students' occupation with self-efficacy related thoughts while reading.

### *Sociometric Rating Score, Self-Concept, and Attitude Towards Reading*

In addition to the discussion of the significant experimental intervention effects on students' reading comprehension and fluency achievement, reportage of strategy use, and self-efficacy related thoughts, it finally should be noticed that no significant findings were discovered in the field of second or fifth graders' sociometric rating score, self-concept, as measured by a Dutch version of Harter's Self-Perception Profile for Children, and attitude towards reading. However, during staff meetings teachers fairly often reported positive evolutions in these areas. Therefore, it can be assumed that the measurement instruments were not apt to assess the changes perceived by the class teachers. On the other hand, taken into account other studies that neither revealed significant outcomes on attitude towards reading (De Corte et al., 2001), self-concept of tutors (Cohen et al., 1982; Griffin & Griffin, 1997) and tutees (Cohen et al., 1982; Franca & Kerr, 1990; Griffin & Griffin, 1997) and that only showed minimal changes on sociometric measures for both tutors and tutees (Franca & Kerr, 1990), one should bear in mind that substantially altering these aspects can not be easily achieved. In this respect, it can be hypothesized that extending and prolonging the experimental treatments will be necessary to promote changes on these rather stable constructs.

### *Consistency of the Results Across Both Studies*

The idea behind replicating the first study in a successive school year was that teachers presumably needed an apprentice period before they would fully master the implementation of the innovative approach towards reading comprehension instruction in their regular instructional practice. In this respect, it was reasonable to think that the results of the replication in the second study would be better than the first study's findings. For some treatments this assumption can be confirmed with regard to second graders' reading comprehension achievement, reported use of strategies, and self-efficacy related thoughts. However, important consistencies throughout both studies arose as well. At first, in neither study significant effects of the experimental treatments were attained on students' sociometric rating score, self-concept, as measured by a Dutch version of Harter's Self-Perception Profile for Children, and attitude towards reading. Secondly, it appeared that the results of the experimental cross-age peer tutoring treatment were reasonably consistent with regard to fifth graders' reading comprehension achievement. Particularly noticeable in this respect are the significant long-term effects, which are even larger than the effects immediately at the end of the treatment. Finally, a negative consistency could be registered, that is the lack of a convincing effect of the reciprocal same-age peer tutoring condition on second and fifth graders' reading comprehension achievement.

### *Discussion with regard to the Second Research Question*

The second research question regarding the surplus value of practicing strategic reading in tutoring dyads in comparison with practice in more traditional teacher-led

activities was addressed by the first study with three experimental groups, among which the strategies-only condition, characterized by explicit reading strategies instruction and teacher-led practice in whole-class activities. Contrary to the expectations based on the literature stressing the importance of peer-led interaction on structured reading activities (e.g., Almasi, 1996; Fuchs & Fuchs, 2000; Fuchs et al., 1997; Greenwood, Delquadri, & Hall, 1989; Klingner et al., 1998; Rosenshine & Meister, 1994), this second question can not be answered positively. Neither for second, nor for fifth graders statistically significant differences were found between the strategies-only condition on the one hand and the reciprocal same-age or cross-age peer tutoring condition on the other hand. Although the differences were not significant, attention should, however, be drawn to the fact that for some measures no significant results were found for the strategies-only condition in comparison with the control condition, while the experimental same-age or cross-age tutoring condition actually did yield significant effects.

Unfortunately, the findings with regard to the second research question can not be compared with previous research results. Most studies investigating the impact of implementing peer tutoring or peer-led interaction about texts equate the innovative instructional technique or approach to a control condition, characterized by traditional instruction without any form of experimental treatment. In this respect, the analyses within the framework of the first research question rather run analogously with the analyses executed in earlier studies. Conversely, to examine the second research question different experimental treatments were mutually compared. Taken into account the literature indicating positive effects of explicit reading strategies instruction in itself (e.g., Baumann et al., 1992; Block, 1993; Dole et al., 1996; Duffy et al., 1987; Haller et al., 1988), it is not completely surprising that no statistically significant surplus value of tutoring activities were found in comparison with the teacher-led practice of reading strategies.

### *Discussion with regard to the Third Research Question*

Unlike most other studies exploring the effects of peer tutoring, the present studies' design allowed the examination of the third research question. More specifically, it was investigated whether the effects of explicit strategies instruction followed by practice in cross-age peer tutoring dyads are different in comparison with identical instruction followed by practice in reciprocal same-age dyads. This design is unique with respect to other studies, which in essence focus on only one tutoring variant. So far, there was only indirect, though nonsignificant, evidence that cross-age tutoring is more effective than same-age tutoring, for the meta-analysis of Cohen et al. (1982) revealed larger effect sizes for both tutors' and tutees' achievement in cross-age tutoring programs. Taken into account the indication from this meta-analysis, it was hypothesized that the impact of the experimental treatments would be more favorable for second and fifth graders respectively functioning as tutees and tutors in cross-age peer tutoring activities than for their peers alternating between the tutor and tutee roles in reciprocal same-age activities.

Generally it can be stated that the pairwise comparisons of the experimental same-age and cross-age tutoring conditions did not reveal statistically significant mutual differences. Three exceptions to this global finding could, however, be registered. Firstly, it should be noticed that opposite to the postulated hypothesis, the second study indicated that at the end of fifth grade children engaged in reciprocal same-age tutoring activities were significantly less occupied with negative self-efficacy related thoughts than students in the cross-age peer tutoring condition. Contrary, the recorded significant differences with regard to second graders' reported strategy use at posttest in the second study and fifth graders' retention test reading comprehension achievement in the first study confirm the surplus value hypothesis of the cross-age tutoring condition. Moreover, when setting the significant and nonsignificant effects side by side and comparing the effect sizes of both conditions, some other research findings also point in this direction. For second graders' reading comprehension and fluency achievement, for example, the results solely revealed a significant effect for the cross-age peer tutoring condition. As considered more extensively above in the discussion of the first research question, this finding can be attributed to the combination of individualized practice and assistance and the competent support of fifth-grade tutors, of which the latter was presumably not provided in second-grade reciprocal same-age dyads. Taken into account the consistency through both studies, the magnitude of the effect sizes, and the long-term effects, it can be concluded that fifth graders also seem to profit more from participating in cross-age peer tutoring dyads than in reciprocal same-age pairs. With regard to accounting for this finding, a reference was made to the fact that practicing strategic reading while functioning as a tutor for a younger student can be seen as a powerful learning environment for enhancing the independent use of reading strategies and for acquiring skills in metacognitive monitoring and regulation of one's own reading process, leading to an improvement in reading comprehension achievement. Moreover, it should be mentioned that the significant long-term results, which were established in both studies, indicate a continuing and even reinforcing cross-age tutoring effect. In sum, it can be concluded that there are some indications that confirm the hypothesis of the surplus value of cross-age tutoring for both tutors and tutees in comparison with reciprocal same-age tutoring.

#### *Discussion with regard to the Fourth Research Question*

The second study allowed us to address the fourth research question and to compare the effectiveness of two variants of teacher training courses developed to gear reading comprehension instruction in elementary schools to the findings and research-based practices emanated from empirical intervention research. More particularly, a one year coaching approach of the teachers was equated to a more restricted 13-hours in-service course. The comparison focused on teachers' experiences with the training course and the actual implementation of the experimental treatments and on student outcomes as a result of the renewed instructional practice.

Both teacher training conditions were designed on the basis of research-based effective training components (Djalil & Anderson, 1989; Harchik, Sherman, Sheldon, &

Strouse, 1992; Showers, 1990; Veenman, Van Tulder, & Voeten, 1994; Wheldall, Merrett, & Borg, 1985), but distinguished themselves mainly by the amount of support lent to the teachers. More specifically, the restricted training was characterized by less counseling and support. However, the following components were included in any case: focused presentation of information about the intended innovations, identification and discussion of the renewed instructional practice, demonstration of relevant instructional behaviors to be implemented, and follow-up booster sessions with in-class supervised practice, on-site consultation, and performance feedback to the teachers. Taking into account the built-in common research-based training features, it was hypothesized that both teacher training strategies would be equally effective.

Generally, the analyses of the experiences reported by the teachers, as well as the student outcomes confirm the hypothesis and corroborate the value of the aforementioned research-based practices included in the compact in-service teacher training. However, in interpreting the findings one has to bear in mind that the teachers of both conditions entered training with different initial experiences, but they were all already experienced in implementing at least part of the interventions, which might have affected the results. In the discussion below, we will go more deeply into this.

With regard to teachers' experiences, teachers in both training conditions reported a strong satisfaction in attending the training course, only little to moderate additional workload because of the implementation of the innovative reading comprehension instruction, and moderate to strong student progress with regard to cognitive, as well as social and emotional development. No large differences between both teacher training conditions were found with regard to teachers' experiences. This finding could possibly be attributed to the fact that at the time of the study, teachers participated for the second consecutive school year. In the previous year teachers in the intensive coaching condition had received a similar elaborated coaching concerning explicit reading strategies instruction and peer tutoring activities. The teachers attending the restricted teacher training course had been involved in an in-service training directed at introducing explicit reading strategies instruction and practicing strategic reading in teacher-led whole-class activities. In this respect, teachers from both conditions were already experienced in implementing at least part of the experimental treatments, which could have counteracted the differences between both teacher training conditions. For only three of the eighteen components in the teacher questionnaire significant differences between both training conditions could be determined. Moreover, two differences favored the restricted in-service course. At first, teachers in the compact training considered the organization of the peer tutoring activities more strongly as an effective opportunity to build in and increase differentiation according to pace or content for children of different abilities. In addition, especially fifth-grade teachers in this condition reported larger student reading comprehension progress than the intensively coached teachers. Presumably, these results can be attributed to the novelty of the peer tutoring activities for the teachers in the restricted in-service condition, whereas teachers in the intensive coaching condition implemented the entire experimental treatment (i.e., explicit reading strategies instruction and peer tutoring activities) for the second consecutive year, which could have led to a habituation effect. It is possible that the former teachers' enthusiasm for the innovative

instructional technique led them to perceive greater reading comprehension progress and may have caused them to appreciate some of the pedagogical advantages of peer tutoring in greater extent. Solely with regard to teachers' perceived workload as a result of the implementation of the experimental treatments, a significant difference favored the intensive coaching condition. It was found that teachers attending the restricted in-service training course experienced a higher workload with regard to settling in the innovative approach and preparing the students for practicing the application of reading strategies in peer tutoring dyads than their colleagues who were intensively coached. Apparently, this finding goes together with the fact that in the restricted in-service teacher training only three three-hours preparatory meetings with the teachers were organized, while in the intensive coaching condition eight local meetings took place, whereby the former group of teachers needed more additional time, apart from the in-service training sessions, to study the intended innovative instructional approach. This reasoning implies that the rather large number of preparatory meetings for the intensively coached teachers were not experienced as overburdening, while the time spent individually on exploring the innovative instructional approach and preparing the students was considered more strongly as an additional burden. The additional workload, as perceived by teachers in the restricted in-service training does, however, not undermine teachers' satisfaction with attending the training course.

As concerns student outcomes it can be concluded that both teacher training conditions were equally effective in changing second and fifth graders' reading comprehension, reading fluency, use of reading strategies, and self-efficacy perceptions towards reading from the start to the end of respectively second and fifth grade, and even till six months after the end of the experimental treatments.

### **Limitations of the Present Studies and Suggestions for Subsequent Research**

As discussed at length in Chapter 2 to 6, the studies in this dissertation, the design, and the applied measurement instruments exhibit some limitations. Moreover, some aspects of the research findings need to be corroborated with further research and evoke new research questions. In this respect, the present section recapitulates briefly the major drawbacks and the aspects that need further attention.

As already quoted in the introductory section of this chapter, the design of the studies enabled us to explore the surplus value of peer tutoring in comparison with teacher-led practice of strategic reading. Moreover, unique with respect to other peer tutoring studies, the effects of cross-age and reciprocal same-age tutoring were contrasted. Taken into account the research findings, it can be argued that explicit reading strategies instruction plays a key role. In addition, practice in cross-age peer tutoring dyads appears to be most favorable for both second and fifth graders.

Notwithstanding the fact that a start was made with comparing the effectiveness of different kinds of reading strategies' practice on student outcomes, a first restriction of the studies, recurrent through the previous chapters, can be mentioned from a methodological point of view. More particularly, it can be argued that due to the



complexity of the experimental treatments, the design did not allow to draw strong conclusions about the relative contribution of the different constituent components of the treatments (e.g., the focus on the integrated teaching and acquisition of a series of different strategies, the instructional approach focusing on explicit teacher explanations, modeling, and a gradual transfer from teacher to student regulation, the regular opportunities to practice strategic reading, ...). Taken into account former studies referred to in Chapter 2, 3, 4, and 6, it can be assumed that it is the combination of different elements that makes a valuable contribution to the positive effects found. However, this is only an assumption. So far, no conclusive evidence was established to explain the efficacy of the experimental treatments. A convincing explanation of what makes the experimental treatments tick, requires a different study. Future research should focus on designing a detailed component analysis to obviate this drawback and to clarify more precisely the essential and perhaps dispensable components of the interventions. This conclusion may also extend to the usefulness of each of the six reading strategies and raises the question whether all of them are equally important. It appears that the fifth reading strategy with regard to monitoring and regulating comprehension is the most central one. The question whether it can be understood and implemented that well without the previous strategies remains however unanswered. In addition, the question arises whether the sixth strategy with regard to classifying types of text really contributes to the learning gains made by the tutor.

A second comment can be made on the fact that for some dependent variables one or more experimental conditions were not equivalent to the control condition on the pretest measure, which imports a threat to the internal validity of the results. Notwithstanding the fact that it is not easy to overcome this kind of problems in quasi-experimental research designs, the recorded significant pretest differences were however quite unexpected, since an effort was made to match teachers and classes in the different research conditions, based on data from the school year preceding the experimental treatment. Anyhow, further research, avoiding internal validity problems, is necessary to verify the present results.

A third shortcoming of the studies is also closely connected to the fact that the experimental treatments were implemented in ecologically valid settings, making it difficult to exercise complete supervision over the total amount of time teachers spent on reading comprehension instruction and practice. In this respect, no differences between the different research conditions were intended. Therefore, the experimental treatments were not meant as an additional program on top of teachers' traditional reading comprehension classes. Experimental teachers were coached to substitute their traditional way of teaching reading comprehension with the innovative instructional practice and they were encouraged to implement the treatments during time normally allocated for reading instruction. To allow us to control for the variable "time spent on reading comprehension" in the analyses, both experimental and control teachers were asked to pass information on the total amount of time they spent on reading comprehension instruction on a weekly and yearly basis. In this respect, all teachers relayed information about the fixed scheduled reading comprehension instruction periods. However, no information was supplied on reading activities apart from these

planned teaching periods. The data teachers provided revealed that in all conditions commensurable amounts of time was allocated to reading comprehension instruction. However, rather large differences were found between the classes within conditions, with some classes spending more than four times as much time on reading comprehension instruction than other classes. Surprisingly, no evidence was found that the class-level total amount of time spent on reading comprehension instruction significantly influenced second or fifth graders' reading comprehension achievement or one of the other response variables. Presumably, this rather unexpected finding can be attributed to the fact that we were not able to catch an overall and accurate picture of the total amount of time teachers spent on reading comprehension instruction and practice. Taken this shortcoming into account, subsequent research should try to assess this potential explanatory variable more precisely, for example by means of observations in different subjects or teachers' logs.

A fourth comment can be made on the fact that the connection between the different outcome variables was not investigated by exploring causal paths. However, two reasons for this shortcoming can be mentioned. The first has to do with a limitation of the actual data analysis techniques. More particularly, the kind of analysis we refer to within the framework of the present studies, requires a combination of both multilevel and structural equation modeling techniques, which is still in the making.

Secondly, even if the necessary techniques would already be fully available, the present findings generally do not give cause to focus on the exploration of the potential links between the outcome variables. However, on the basis of the second study's results, two exceptions can be made in this respect. More particularly, it can be assumed that the effect on second graders' reported reading strategy use mediates students' significant reading comprehension progress. The same mediating function can be attributed to the effect on second graders' positive self-efficacy related thoughts. However, with regard to this variable it can just as well be hypothesized that the significant result is an interesting side effect of the experimental treatment. Future research should endeavor to elucidate these assumptions. In this respect, some first explorative analyses were performed by including students' reported strategy use and self-efficacy related thoughts in the second study's multilevel model with regard to second graders' reading comprehension achievement. The fact that the significant posttest effect of the cross-age peer tutoring condition did not drop by including these variables, does not point in the direction of mediating effects.

A fifth major restriction of the studies has to do with the measurement instrument used to assess students' application of reading comprehension strategies. Since explicit reading strategies instruction as well as organizing peer tutoring activities to put these strategies into practice are involved with assisting students with the process of comprehending text, measuring students' use of reading strategies deservedly merited attention in both studies. To that end a self-report scale was developed. Especially for fifth graders, however, the findings with regard to this questionnaire were not consistent with the experimental intervention results revealed on general reading comprehension achievement. It seems quite reasonable, however, that the positive effects of the experimental interventions on fifth graders' reading comprehension

achievement can be attributed to the ameliorated metacognitive monitoring and regulation of one's reading process. In addition it can be hypothesized that the reading strategies, practiced by means of the structured strategy assignment cards, act as tools in these metacognitive processes. Nevertheless, there is no conclusive evidence in this respect, since no significant effects were recorded in the field of fifth graders' reports of reading strategy use.

As already mentioned above, it can be hypothesized that the nonsignificant results with regard to the reported reading strategy use can be attributed to the fact that the questionnaire was not an appropriate measurement instrument for this age group. More specifically, it can be assumed that there is a tendency to social desirability in answering the questionnaire, for the use of a self-report questionnaire has the disadvantage of being an indirect measurement, which entails that children's reportage of strategy use is not necessarily consistent with their actual behavior while reading (Veenman & van Hout-Wolters, 2002). In this respect, future research should substitute the use of the questionnaire by more objective measures. Finding reliable and valid measurement instruments to assess students' reading behavior, and application of reading strategies is however not easy. Even more difficult to assess, especially with elementary school students, is the metacognitive monitoring and regulation of the reading process. As Topping (2001) argues, there is even no right way to do it. Therefore, using more than one outcome measure is a wise strategy. In this respect, thinking-aloud protocols, behavior observations, and retrospective techniques, like stimulated recall, can offer opportunities to shed light on students' generalization of reading strategy use, internal cognitive processes, and metacognitive functioning in different formats and contexts.

Thinking-aloud protocols, which are the verbal reports produced by subjects who express their thoughts while engaged in some activity (Kucan & Beck, 1997), have the advantage of an on-line registration of one's cognitive and metacognitive processes (Veenman & van Hout-Wolters, 2002). However, within the framework of the present studies characterized by the implementation of peer tutoring activities, stimulated recall techniques, such as applied by Almasi et al. (1996) and Alvermann et al. (1996), might be a better alternative to prompt students to reflect on and express their thinking during peer-led interactions about texts. Having tutors think aloud during peer tutoring sessions is not obvious since tutees are reading aloud already. But, when viewing videotaped fragments of peer tutoring discussions in which they participated, tutors can be asked to reflect on their use of the strategy assignment cards, on the types of questions they ask and the assistance they provide, and on the timing of their questions and helping behavior. On the other hand, the stimulated recall for tutees can focus on the reasons of their calls for help or mistakes, and on their understanding of the underlying aim of the strategy cards and tutors' questions. Thinking-aloud protocols could of course be used apart from the peer tutoring situation when testing the tutors' progress in reading achievement when reading texts on their own reading level. This could allow us to investigate the hypothesized transfer of the metacognitive processes involved in monitoring and regulating a tutees' reading to the metacognitive monitoring and regulation of one's own reading process.

In addition to the shortcoming of the self-report scale with regard to students' reading strategy use, some other restrictions with regard to the measurement occasions and

instruments used to investigate the effects of the experimental interventions on student outcomes should be mentioned. With regard to the measurement occasions, the studies were confined to a pre-, post-, and retention test, which respectively took place at the start of the school year (i.e., before the experimental interventions), at the end of the school year (i.e., at the end of the interventions), and six months after the end of the interventions. In this respect it can be argued that including interim measures would yield additional, relevant information with regard to the interventions' short-term effects and students' evolution throughout the school year.

As concerns the outcome measures, some instruments are assumed to give rise to an incomplete or presumably unsuitable assessment of the intended student skill or characteristic. In reference to the applied standardized reading fluency test, for example, the remark is passed that the test solely assessed accuracy and rate of words read in isolation, while subsequent research should focus on context-bound texts as well. On the other hand, as concerns the nonsignificant findings with regard to students' sociometric rating score, self-concept, and attitude towards reading, it can be hypothesized that the applied instruments were not apt to assess student changes, that were however perceived by the class teachers.

In general, it can be concluded that most restrictions with regard to the measurement instruments used can be obviated by deploying multiple outcome measures to triangulate the data collection.

Finally, it can be argued that it is important to gather information about the extent to which the implementation of the experimental treatments has been carried out correctly and in accordance with the intention of the teacher training program. Notwithstanding the fact that regular discussions with the experimental teachers took place and observations in the classes were executed - especially with the teachers in the intensive coaching condition -, a comment can be made on the fact that quality of implementation data were not collected systematically. In addition, no in-depth analysis of the reading activities in the control classes was performed. Although the interviews with the control group teachers suggested that little or no attention was paid to the intentional and systematic teaching of reading strategies and no peer tutoring or other forms of peer-led interaction were used, it was practically impossible within the framework of the present studies to collect systematic data about the actual instruction in all the control classes. Therefore, subsequent research should strive to collect good quality data on the characteristics of the control group's comprehension instruction on the one hand and on the treatment or implementation integrity of the innovative reading comprehension instruction approach on the other hand. In this respect, the use of a structured checklist by a double-blind observer could be useful.

Besides the recommendations for future research that result from the limitations of the studies in this dissertation, some aspects of the research findings can be supported with more substantial evidence and empirical support as well.

At first, it can be argued that similar studies with larger sample sizes at a wider range of elementary grade levels using a variety of measurement instruments will provide researchers with a greater understanding of the effects of explicit reading strategies

instruction and cross-age or reciprocal same-age peer tutoring activities on students' reading skills and social and emotional development.

Secondly, the fact that long-lasting treatment effects on student outcomes were solely found in some cases leads us to suspect that long-term effects can only be attained by continuing the interventions. Subsequent research pursuing the experimental treatments in the following grades should try to elucidate this assumption, especially with a view to determine the prerequisites for developing independent and self-regulative readers for life. In addition, the assessment of transfer effects is a critical area for further research as well. More information is needed about how often and how long treatments must be implemented to promote routine use and transfer of the skills to other assignments or fields of learning.

In the third place, it should be noticed that notwithstanding the fact that pairwise mutual comparisons of the experimental conditions' effects on student outcomes rarely revealed statistically significant contrasts, noticeable differences occurred when setting the significant and nonsignificant effects side by side and comparing the effect sizes of the experimental conditions. Future, more qualitatively oriented systematic research should try to elucidate the differential effects of practicing strategic reading in teacher-led, reciprocal same-age, or cross-age tutoring conditions. In this respect, detailed in-depth observation and analysis of the content, the nature, and quality of students' interactions in either teacher-led whole-class discussions, cross-age, or reciprocal same-age dyads will be requisite. Presumably, small-scale studies, with frequent and elaborated observations will yield high quality data.

Further, the unresolved question with regard to the contradictory results depending on the data analysis technique used (i.e., separate two-level regression analyses on students' post- and retention test data versus a combined three-level repeated measures model) merits particular attention. In addition, a four-level model including the dyads students cooperated in as a separate level between the class-level and the student-level deserves exploration as well.

Finally, further research is necessary to investigate whether the quality of the in-service teacher training program and the connected teacher satisfaction can be equaled by disseminating the innovative instructional approach to reading comprehension instruction on a broad scale through regular in-service training authorities.

## **Conclusion**

Notwithstanding the reported restrictions and the fact that further research into the effectiveness of blending explicit reading strategies instruction and peer tutoring is called for, the presented results of the studies lend additional support to the cognitively based views of the reading comprehension process. These underline that expert readers distinguish themselves from novice or poor readers by their disposition of metacognitive skills to monitor and regulate comprehension and by the use of a flexible repertoire of cognitive "fix up" strategies to repair understanding when it breaks down. In this respect, a new model of comprehension instruction is suggested, in which explicit instruction in a set of reading strategies that characterize proficient readers is emphasized, as well as promoting students' understanding about when and

how to use these strategies in a flexible way. In addition also the application of instructional techniques that challenge students to actively engage in negotiating and constructing meaning from texts is stressed. In that way, students are supported in their process of gradually emerging expertise.

The findings within the framework of this dissertation demonstrate the educational benefits of composing reading comprehension instruction from the cognitively based models of the comprehension process and give support to the need to modify the prevailing reading comprehension curriculum and instructional methods. More particularly, the outcomes of the studies generally corroborate the effectiveness of explicit reading strategies instruction followed by practice in peer tutoring or teacher-led activities as a feasible tool to promote regular elementary students' reading comprehension achievement. Consequently, the implemented experimental treatments may merit a place in the regular curriculum and teaching methods. In contrast with other studies, a key characteristic of the present studies is the explicit comparison of the effectiveness of reciprocal same-age and cross-age tutoring. In this respect, an important finding is that in comparison with the other research conditions the cross-age peer tutoring condition generally revealed the most favorable results for both second-grade tutees and fifth-grade tutors. A conceivable interpretation of this finding refers to the nature of the tutoring task and the accompanying responsibility tutors accept when supporting other students. Peer tutoring requires a variety of different skills from the tutors. They need to establish and maintain a positive relationship with their tutee by showing interest, supporting, complimenting, encouraging, and praising them. They are expected to scaffold the tutees' learning by superintending their reading process attentively, by making explicit the metacognitive processes involved in reading, by modeling and imparting the selection, application, and evaluation of relevant reading strategies, and by negotiating the meaning of their reading. They have to judge what questions to ask and when to ask them, to correct errors, to provide appropriate explanations, help, assistance, and feedback that can be useful to the tutee, and to offer reinforcement. This all entails that tutors themselves are continuously challenged to consider the subject fully from different perspectives, to engage in active monitoring and regulation of their tutees' reading, to reorganize and clarify their own knowledge and understandings, and to elaborate on information. Presumably, meeting this combination of tasks and responsibilities is easier when working with students on a lower reading level than one's own, such as in the case of fifth-grade tutors reading with second graders in cross-age peer tutoring activities instead of reading with a classmate in same-age tutoring activities. Hence, it appears that practicing the use of reading strategies while functioning as a tutor for a younger student functioning on a lower achievement level, creates the best opportunities for fifth graders to give full attention to practice metacognitive monitoring and regulation skills and relevant reading strategy selection and application themselves, which could explain their reported most favorable and persistent learning gains in the field of comprehension. On the other hand, it can be assumed that fifth graders in comparison with second graders are more mature and better capable of performing the tutor role and the tasks and responsibilities that go with it, which can account for the finding that second graders led by a fifth-grade tutor come off better than second graders alternating regularly between the roles of tutor and tutee.

Notwithstanding the reported positive results of explicit reading strategies instruction followed by practice in peer tutoring or teacher-led activities, we do not aver that after a year of strategic reading instruction students have become expert self-regulated readers and learners. This is especially demonstrated by the fact that solely for some fifth-grade conditions long-term effects of the experimental treatments on students' comprehension achievement could be established. Following Brown et al. (1996) our hypothesis is that "true self-regulation is the product of years of literacy experiences" (p. 34). However, the developed innovative instructional approaches, which were successfully implemented in naturally constituted classes as part of the overall curriculum, in particular explicit strategy instruction in conjunction with cross-age peer tutoring, show considerable promise to get this process off to a good start.

In addition to the reported effectiveness of the experimental treatments on students' reading comprehension, the findings of the studies carry significant implications for reading fluency, strategy use, and self-efficacy related thoughts as well. However, subsequent more in-depth research is necessary to clear the lack of univocal results for second and fifth graders, to gain insight in students' underlying metacognitive processes, and to explore methods of maximizing explicit strategy instruction's and peer tutoring's potential in this respect.

Finally, the results underline that the innovative approach to reading comprehension instruction which was developed in the present study can be successfully implemented in daily instruction and practice in regular elementary school classes, even by means of a relatively concise in-service teacher training, about which teachers generally are well satisfied. More specifically, the findings confirm the efficacy of in-service training embracing a focused presentation of information about the background of the innovative instructional approach; an identification and discussion of the intended instructional practice; a feasible, elaborated and easy-reference manual; a demonstration of relevant instructional behaviors to be implemented by means of accurately selected video fragments; and follow-up sessions with in-class supervised practice, on-site consultation, and performance feedback to the teachers.

## References

- Aarnoutse, C. A. J. (1996). *Leesattitudeschaal* [Reading attitude scale]. Nijmegen, The Netherlands: Berkhout.
- Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal*, 35, 309-332.
- Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.
- Almasi, J. F., McKeown, M. G., & Beck, I. L. (1996). The nature of engaged reading in classroom discussions of literature. *Journal of Literacy Research*, 28, 107-146.
- Alvermann, D. E., Young, J. P., Weaver, D., Hinchman, K. A., Moore, D. W., Phelps, S. F., Thrash, E. C., & Zalewski, P. (1996). Middle and high school students'

- perceptions of how they experience text-based discussions: A multicase study. *Reading Research Quarterly*, 31, 244-267.
- Bast, J. W., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, 34, 1373-1399.
- Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior: A Journal of Literacy*, 24, 143-172.
- Block, C. C. (1993). Strategy instruction in a literature-based reading program. *The Elementary School Journal*, 94, 139-151.
- Brand-Gruwel, S., Aarnoutse, C. A. J., & Van Den Bos, K. P. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and instruction*, 8, 63-81.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.
- De Corte, E., Verschaffel, L., & Van De Ven, A. (2001). Improving text comprehension strategies in upper primary school children: A design experiment. *British Journal of Educational Psychology*, 71, 531-559.
- Dixon-Kraus, L. A. (1995). Partner reading and writing: Peer social dialogue and the zone of proximal development. *Journal of Reading Behavior*, 27, 45-64.
- Djalil, A., & Anderson, L. W. (1989). The impact of a research-based teacher training program on Indonesian teachers, classrooms and students. *Teaching and Teacher Education*, 5, 165-178.
- Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly*, 31, 62-88.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliff, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putnam, J., & Bassari, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22, 347-368.
- Franca, V. M., Kerr, M. M. (1990). Peer tutoring among behaviorally disordered students: Academic and social benefits to tutor and tutee. *Education and Treatment of Children*, 13, 109-129.
- Fuchs, L. S., & Fuchs, D. (2000). Building student capacity to work productively during peer-assisted reading activities. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 95-115). New York: Teachers College Press.
- Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.



- Griffin, B. W., & Griffin, M. M. (1997). The effects of reciprocal peer tutoring in graduate students' achievement, test anxiety, and academic self-efficacy. *The Journal of Experimental Education*, 65, 197-209.
- Haller, E. P., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher*, 17, 5-8.
- Harchik, A. E., Sherman, J. A., Sheldon, J.B., & Strouse, M.C. (1992). Ongoing consultation as a method of improving performance of staff members in a group home. *Journal of Applied Behavior Analysis*, 25, 599-610.
- Harter, S. (1985). *Manual for the self-perception profile for children*. Denver: University of Denver.
- Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: verbal versus visual strategies. *Journal of Educational Psychology*, 92, 772-782.
- Jones, K., & Duncan, C. (1998). Modeling context and heterogeneity: Applying multilevel models. In E. Scarbrough, & E. Tanenbaum (Eds.), *Research strategies in the social sciences. A guide to new approaches* (pp.95-123). Oxford, United Kingdom: Oxford University Press.
- Kelly, M., & Moore, D. W. (1994). Reciprocal teaching in a regular primary school classroom. *Journal of Educational Research*, 88, 53-62.
- Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, 96, 275-293.
- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, 99, 3-21.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271-299.
- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Paterson, L., & Goldstein, H. (1991). New statistical methods for analyzing social structures: An introduction to multilevel models. *British Educational Research Journal*, 17, 387-393.
- Pressley, M. (2000). Comprehension instruction in elementary school: A quarter-century of research progress. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.), *Reading for meaning. Fostering comprehension in the middle grades* (pp. 32-51). New York: Teachers College Press.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92, 513-555.

- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal*, 90, 3-32.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., & Draper, D. (1999). *A user's guide to MLwiN*. London: Institute of Education.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, 64, 479-530.
- Showers, B. (1990). Aiming for superior classroom instruction for all children: A comprehensive staff development model. *Remedial and Special Education*, 11, 35-39.
- Simmons, D. C., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice*, 9, 203-212.
- Simmons, D. C., Fuchs, L. S., Fuchs, D., Mathes, P. G., Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. *The Elementary School Journal*, 95, 387-408.
- Singleton, L. C., & Asher, S. R. (1977). Peer preferences and social interaction among third-grade children in an integrated school district. *Journal of Educational Psychology*, 69, 330-336.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced modeling*. London: Sage.
- Sticht, T. G., & James, J. H. (1984). Listening and reading. In P. D. Pearson (Ed.), *Handbook of reading research* (Vol. 1, pp. 293-318). New York: Longman.
- Topping, K. J. (2001). *Thinking Reading Writing. A practical guide to paired learning with peers, parents and volunteers*. London: Continuum.
- Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2002). *Het meten van metacognitive vaardigheden* [Measuring metacognitive skills]. Paper presented at the ORD-conference, May 30<sup>th</sup>, 2002, Antwerp.
- Veenman, S., Van Tulder, M., & Voeten, M. (1994). The impact of in-service training on teacher behavior. *Teaching and Teacher Education*, 10, 303-317.
- Veerman, J. W., Straathof, M. A. E., Treffers, Ph. D. A., Van den Bergh, B. R. H., & ten Brink, L. T. (1997). *Competentiebelevingsschaal voor Kinderen (CBSK)* [Self-Perception Profile for Children]. Lisse, The Netherlands: Swets & Zeitlinger.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walraven, M. (1995). *Instructie in leesstrategieën. Problemen met begrijpend lezen en het effect van instructie aan zwakke lezers* [Instruction in reading strategies. Difficulties with reading comprehension and the effect of instruction for poor readers]. Amsterdam: Paedologisch Instituut.
- Wheldall, K., Merrett, F., & Borg, M. (1985). The behavioural approach to teaching package (BATPACK): An experimental evaluation. *British Journal of Educational Psychology*, 55, 65-75.